DOCUMENT RESUME

ED 346 107 TM 018 202

| | |
|---|---|
| AUTHOR | Sanders, Petrus Franciscus |
| TITLE | The Optimization of Decision Studies in Generalizability Theory. |
| INSTITUTION | Centraal Inst. voor Toetsonwikkeling, Arnhem (Netherlands). |
| REPORT NO | ISBN-90-9004906-1 |
| PUB DATE | 92 |
| NOTE | 112p.; Master's Thesis, University of Amsterdam. Chapters 2 through 6 were previously published in several periodicals. |
| PUB TYPE | Dissertations/Theses - Doctoral Dissertations (041) |
| EDRS PRICE | MF01/PC05 Plus Postage. |
| DESCRIPTORS | Budgeting; Data Interpretation; *Decision Making; Equations (Mathematics); Foreign Countries; *Generalizability Theory; Mathematical Models; *Measurement Techniques; *Research Design; Retrenchment; *Test Construction |
| IDENTIFIERS | Decision Analysis; *Mathematical Programing; *Optimization |

ABSTRACT

The application of mathematical programming techniques is extended to the construction of measurement instruments in generalizability theory. Key concepts in generalizability theory are explained and a description is given of: (1) the one-facet crossed design; (2) the two-facet crossed design; and (3) a two-facet nested design. The optimization of decision studies is also discussed. The current study was undertaken to provide test constructors with procedures to determine the optimal design for measurement instruments with more than one facet. Procedures were developed for two pertinent test construction problems. Chapter 1 is an introduction. Chapter 2 describes the minimization of the number of observations given a minimum acceptable generalizability coefficient. Chapter 3 presents a procedure for the maximization of the generalizability coefficient of a measurement instrument given limited resources. An additional two-step procedure for minimizing the generalizability coefficient under a budget constraint is given in Chapter 4. Chapter 5 offers alternative solutions to maximizing the generalizability coefficient under budget constraints. Chapter 6 addresses the problem of sampling variability of variance components in generalizability theory. The text is in English and a Dutch foreword and abstract are included. Sixty-seven references are included. (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P.F. SANDERS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

CIP-DATA KONINKLIJKE BILBIOTHEEK, DEN HAAG

Sanders, Petrus Franciscus

The optimization of decision studies in generalizability theory / Petrus Franciscus Sanders. - [S.l. : s.n.]. - Ill.
Thesis Universiteit van Amsterdam. - With ref. - With summary in Dutch.
ISBN 90-9004906-1
Subject heading: mathematical programming.

© Copyright 1992 P.F. Sanders, National Institute for Educational Measurement (Cito), Arnhem.
No part of this book may be reproduced in any form, by print, photoprint, microfilm or any means without written permission from the author.

# The Optimization of Decision Studies in Generalizability Theory

**ACADEMISCH PROEFSCHRIFT**

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam,
op gezag van de Rector Magnificus
prof. dr. P.W.M. de Meijer
in het openbaar te verdedigen in de Aula der Universiteit
(Oude Lutherse Kerk, ingang Singel 411, hoek Spui),
op donderdag 2 april 1992 te 13.00 uur

door Petrus Franciscus Sanders

geboren te Wijchen

Promotor:
Prof.dr. G.J. Mellenbergh

Co-promotor:
Dr. D.N.M. de Gruijter
(Rijksuniversiteit Leiden)

Leden promotiecommissie:
Prof.dr. L.J.Th. van der Kamp
(Rijksuniversiteit Leiden)
Prof.dr. W.J. van der Linden
(Universiteit Twente)
Prof.dr. W. Molenaar
(Rijksuniversiteit Groningen)

Faculteit der Psychologie

Voor mijn ouders,
Johannes en Helena Sanders

## Voorwoord

Op deze plaats wil ik iedereen bedanken die een bijdrage heeft geleverd aan de totstandkoming van dit proefschrift.

In de eerste plaats Phiel Theunissen, Sjoerd Baas en Theo Eggen die mede-auteur(s) zijn van publikaties die in dit proefschrift de hoofdstukken 2, 3 en 4 vormen. Met name wil ik 'buitenstaander' Sjoerd speciaal bedanken voor onze plezierige samenwerking die tot nu toe geresulteerd heeft in twee artikelen.

Mijn promotor, Don Mellenbergh, en co-promotor, Dato de Gruijter, hebben mij bij het schrijven van dit proefschrift uitstekend begeleid. Hun altijd opbouwende kritiek heeft de afronding van het proefschrift aanzienlijk bespoedigd.

De produktie van dit proefschrift was in uitstekende handen bij Betty Herrewijn. Dat het proefschrift er zo fraai uitziet is uitsluitend haar verdienste. Voor de vermenigvuldiging van het proefschrift heeft de drukkerij van het Cito zorg gedragen.

Zonder de hulp van mijn vrouw Marianne zou er in dit proefschrift vele malen tegen de Engelse taal gezondigd zijn. Aan de leesbaarheid van dit proefschrift heeft zij in niet geringe mate bijgedragen.

Mijn werkgever, het Cito, wil ik bedanken voor de faciliteiten die het schrijven van dit proefschrift mogelijk hebben gemaakt. Zouden aan dit proefschrift stellingen toegevoegd zijn, dan zou de laatste stelling zeker 'Het Cito moet blijven' geweest zijn.

Zetten, februari 1992

Piet Sanders

# TABLE OF CONTENTS


Voorwoord
Table of Contents viii

1 Introduction 1
1.1 Introduction 3
1.2 Key Concepts in Generalizability Theory 3
1.2.1 One-Facet Crossed Design 4
1.2.2 Two-Facet Crossed Design 6
1.2.3 Two-Facet Nested Design 7
1.3 The Optimization of Decision Studies 8
1.3.1 Optimization of the One-Facet Random-Model Crossed Design 11
1.3.2 Optimization of the Two-Facet Random-Model Crossed Design 12
1.3.3 Optimization of the Two-Facet Random-Model Nested Design 14
1.4 Purpose of the Study 14
1.4.1 Overview of the Study 15
References 18

2 Minimizing the Number of Observations: A Generalization of the Spearman-Brown Formula 21
2.1 Introduction 23
2.2 Branch-and-Bound Algorithm 27
2.3 Branch-and-Bound for the Two-Facet Design Problem 29
2.4 Branch-and-Bound for the Multi-Facet Design Problem 35
2.5 Conclusions and Discussion 39
References 42

3 Maximizing the Coefficient of Generalizability Under the Constraint of Limited Resources 43
3.1 Introduction 45
3.2 Optimization Problem 47
3.3 Solution for the Two-Facet Design Problem 48
3.3.1 Continuous Solution 48
3.3.2 Integer Solution 52
3.4 Example 54
3.5 Conclusions and Discussion 58
References 62

4 The Optimum Allocation of Measurements in Balanced Random Effects Models 63
4.1 Introduction 65
4.2 Optimization Problem 66
4.3 Proposed Solution Three-Way Design 67
4.3.1 Continuous Solution 67
4.3.2 Integer Solution 68
4.3.3 Example 69
4.4 Proposed Solution Four-Way Design 72
4.5 Conclusions and Discussion 74
References 75

5 Alternative Solutions for Optimization Problems in Generalizability Theory 77
5.1 Introduction 79
5.2 Solutions for Two-Facet Optimization Problems 81
5.3 Two-Facet Example 84
5.4 Conclusions and Discussion 85
References 87

6 Sampling Errors of Variance Components 89
6.1 Introduction 91
6.2 Sampling Errors of Estimates of Variance Components 92
6.3 Minimizing Sampling Errors of Estimates of Variance Components 95
6.3.1 Intuitive Approach 95
6.3.2 Mathematical Programming Approach 98
6.4 Conclusions and Discussion 101
References 103

7 Epilogue 105
References 110

Samenvatting
De optimalisering van decisiestudies in de generaliseerbaarheidstheorie 111

# Chapter 1

# Introduction

## 1.1 Introduction

Although Votaw (1952) was the first to introduce mathematical programming to test theory, it was an article by Theunissen (1985), however, that marked the beginning of the application of mathematical programming techniques to the construction of tests. Numerous studies (e.g., Boekkooi-Timminga, 1989; Adema, 1990) have since presented new procedures for test construction in classical test theory and item response theory. As a result, test constructors now have at their disposal mathematical procedures to select the items that optimally meet their quantitative and qualitative test specifications, and no longer have to revert to the trial and error procedures (e.g., Gulliksen, 1950; Lord, 1977) employed in the past.

In the present study the application of mathematical programming techniques was extended to the construction of measurement instruments in generalizability theory. Extensive expositions of generalizability theory can be found in monographs by Cronbach, Gleser, Nanda, and Rajaratnam (1972), Brennan (1983), and Cardinet and Tourneur (1985). The key concepts in generalizability theory and a description of three designs, the one-facet crossed design, the two-facet crossed design, and a two-facet nested design, are presented in section 1.2. The optimization of designs in general and the optimization of the three designs mentioned are discussed in section 1.3. The purpose of the study and an overview of Chapters 2 to 6 are presented in section 1.4.

## 1.2 Key Concepts in Generalizability Theory

A common measurement procedure in testing is to have examinees answer questions or items that are graded by raters. Feldt and Brennan (1989, p. 128-129), used this measurement procedure to explain the central concepts in generalizability theory. In generalizability theory the term *population* is used for the *objects of measurement* which in the present context are persons. Items and raters are called measurement

facets. A *facet* is a set of similar *conditions of measurement*. Items and raters together constitute the *universe of admissible observations*. In a *G(oneralizability)-study*, a sample of $n_p$ persons can be administered a sample of $n_i$ items graded by a sample of $n_j$ raters. As each person is observed under every pairing of the conditions of the two facets, the design of this G-study is called a *crossed design*. Given an infinite number of items and raters, an *analysis of variance* of the actual observations will result in estimates of seven *variance components* of the *random-model*. The variance components obtained in the G-study are used to make decisions about the measurement objects in a *D(decision)-study*. In a D-study, test administrators specify the universe of admissible observations to which they want to generalize the actual observations. The *universe of generalization* can consist of all the conditions in the universe of admissible observations or a subset of the conditions. After the universe of generalization is determined, the D-study is completed by estimating the *D-study variance components*, *universe score variance*, *error variance*, and the *generalizability coefficient*.

### 1.2.1 One-Facet Crossed Design

An example of a one-facet crossed design is one in which a number of items are administered to all the persons to be tested. The universe of admissible observations consists of the item facet only. The observed score $X_{pi}$ of person $p$ on item $i$ is expressed by the following linear model as

$$X_{pi} = \mu + \mu_p\sim + \mu_i\sim + \mu_{pi}\sim, \qquad (1)$$

where $\mu$ is the overall grand mean, $\mu_p\sim$ $(= \mu_p - \mu)$ is the person effect, $\mu_i\sim$ $(= \mu_i - \mu)$ is the item effect, and $\mu_{pi}\sim$ $(= X_{pi} - \mu_p - \mu_i + \mu)$ is the residual. The residual is composed of $\mu_{pi}$, an interaction component, and $e_{pi}$, a random error component. With a single observed

score for each person by item combination these components are confounded and therefore also denoted by $\mu_{pi,e}$. Assuming a random model, all effects in (1), except $\mu$, are called random effects because they are associated with a random sampling from the infinite population of persons and the infinite universe of items. For each effect or component in (1), there is an associated variance which is called the variance component. The observed score variance,

$\sigma_x^2 = \sigma^2(X_{pi}) = \mathop{\mathcal{E}}_{p}\mathop{\mathcal{E}}_{i}(X_{pi}-\mu)^2$, can be expressed as

$$\sigma_x^2 = \sigma_p^2 + \sigma_i^2 + \sigma_{pi,e}^2,$$

where $\sigma_p^2$ is the variance component for persons or universe score variance, $\sigma_i^2$ is the variance component for the items facet, and $\sigma_{res}^2 = \sigma_{pi,e}^2$ is the variance component for the persons by items interaction plus error. The generalizability coefficient, denoted by $\mathcal{E}\rho^2$, is defined as

$$\mathcal{E}\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pi,e}^2}. \qquad (2)$$

Equation (2) gives the ratio of universe score variance to the expected observed score variance for a randomly chosen single condition. Since a test contains more than one item, a generalizability coefficient of an average score based on $n$ items is required. This stepped-up generalizability coefficient is defined as

$$\mathcal{E}\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pi,e}^2}{n}}. \qquad (3)$$

The generalizability coefficient in (3) expresses, on a 0-1 scale, how well persons can be differentiated while generalizing over the one facet. The coefficient is approximately equal to the expected value of the squared correlation between observed and universe scores. The

coefficient may also be interpreted as an approximation to the correlation between the scores from two randomly parallel tests. Generalizability coefficient (3) with estimated variance components is identical to Cronbach's coefficient alpha.

### 1.2.2 Two-Facet Crossed Design

For the two-facet crossed design discussed above with facet $I(i=1,\ldots,I)$ being items and facet $J(j=1,\ldots,J)$ raters, the observed score of a person can be decomposed as

$$X_{pij} = \mu + \mu_p\sim + \mu_i\sim + \mu_j\sim + \mu_{pi}\sim + \mu_{pj}\sim + \mu_{ij}\sim + \mu_{pij}\sim, \qquad (4)$$

where $X_{pij}$ is the score awarded to person $p$ on item $i$ by rater $j$, $\mu$ is the overall grand mean, $\mu_p\sim$ is the person effect, $\mu_i\sim$ is the item effect, $\mu_j\sim$ $(= \mu_j - \mu)$ is the rater effect, $\mu_{pi}\sim$ $(= \mu_{pi} - \mu_p - \mu_i + \mu)$ is the person x item effect, $\mu_{pj}\sim$ $(= \mu_{pj} - \mu_p - \mu_i + \mu)$ is the person x rater effect, $\mu_{ij}\sim$ $(= \mu_{ij} - \mu_i - \mu_j + \mu)$ is the item x rater effect, and $\mu_{pij}\sim$ $(= X_{pij} - \mu_{pi} - \mu_{pj} - \mu_{ij} + \mu_p + \mu_i + \mu_j - \mu)$ is the residual effect. The observed score variance, $\sigma^2_X = \sigma^2(X_{pij})$, can be expressed as

$$\sigma^2_X = \sigma^2_p + \sigma^2_i + \sigma^2_j + \sigma^2_{pi} + \sigma^2_{pj} + \sigma^2_{ij} + \sigma^2_{pij,e},$$

where $\sigma^2_p$ is the variance component for persons, $\sigma^2_i$ is the variance component for items, $\sigma^2_j$ is the variance component for raters, $\sigma^2_{pi}$ is the variance component for the persons x items interaction, $\sigma^2_{pj}$ is the variance component for the persons x raters interaction, $\sigma^2_{ij}$ is the variance component for the items x raters interaction, $\sigma^2_{res} = \sigma^2_{pij,e}$ is the variance component for the persons x items x raters interaction plus error.

The stepped-up coefficient for the two-facet random-model crossed design is defined as

$$\mathcal{E}\rho^2 = \frac{\sigma^2_p}{\sigma^2_p + \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{pj}}{n_j} + \frac{\sigma^2_{pij,e}}{n_i n_j}}, \qquad (5)$$

where $n_i$ and $n_j$ denote the number of items and raters, respectively, and $n_i n_j$ denotes the product of the number of items and raters which equals the number of observations per person. The generalizability coefficient in (5) expresses how well persons can be differentiated while generalizing over the item and rater facet. The coefficient can also be interpreted as the expected correlation between scores based on $n_i$ items and $n_j$ raters, and scores based on a different set of $n_i$ items and a different group of $n_j$ raters. Mellenbergh (1977) introduced the concept of *coefficient of replicability* for the correlation coefficient between a particular measure in an experiment and the measure obtained in a particular replication of the experiment. Generalizability coefficients as well as some coefficients derived by Maxwell and Pilliner (1968) are particular elements of this set of replicability coefficients.

#### 1.2.3 Two-Facet Nested Design

An example of a (partial) nested design is one in which the same number of items are administered to all the persons of a sample, but item 1 (i=1) is graded by rater 1 (j=1) and rater 2 (j=2), item 2 (i=2) is graded by rater 3 (j=3) and rater 4 (j=4), etc. For this two-facet nested design, the decomposition of a person's observed score is expressed as

$$X_{pij} = \mu + \mu_p\sim + \mu_i\sim + \mu_{j:i}\sim + \mu_{pi}\sim + \mu_{pj:i}\sim, \qquad (6)$$

where $\mu$ is the overall grand mean, $\mu_p\sim$ is the person effect, $\mu_i\sim$ is the item effect, $\mu_{j:i}\sim$ $(= \mu_{ij} - \mu_i)$ is the rater nested (denoted by a colon) within item effect, $\mu_{pi}\sim$ $(= \mu_{pi} - \mu_p - \mu_i + \mu)$ is the person x item effect, and $\mu_{pi:j}\sim$ $(= X_{pij} - \mu_{pj} - \mu_{ij} + \mu_j)$ is the residual effect. The observed score variance, $\sigma_x^2 = \sigma^2(X_{pij})$, can be expressed as

$\sigma^2_x = \sigma^2_p + \sigma^2_i + \sigma^2_{j,ij} + \sigma^2_{pi} + \sigma^2_{pj,pij,e}$, where $\sigma^2_p$ is the variance component for persons or universe score variance, $\sigma^2_i$ is the variance component for items, $\sigma^2_{j,ij}$ is the variance component for raters nested within items (i.e., confounded sum of $\sigma^2_j$ and $\sigma^2_{ij}$), $\sigma^2_{pi}$ is the variance component for the persons x items interaction, $\sigma^2_{res} = \sigma^2_{pj,pij,e}$ is the variance component for the persons x raters nested within items interaction plus error (i.e., confounded sum of $\sigma^2_{pj}$, $\sigma^2_{pij}$, and $\sigma^2_e$).

The stepped-up generalizability coefficient for the two-facet random-model nested design is defined as

$$\mathcal{E}\rho^2 = \frac{\sigma^2_p}{\sigma^2_p + \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{pj,pij,e}}{n_i n_j}}. \qquad (7)$$

### 1.3 The Optimization of Decision Studies

In the D-study or optimization phase of a generalizability analysis, decisions have to be made on the aim and the composition of the measurement instrument or measurement procedure.

One of the most fundamental modifications is to employ a different design in the D-study than in the G-study. A crossed design can be changed into a (partially) nested facet design by nesting one facet within another facet. The crossed design discussed in section 1.2.2 can be changed into a nested design by having different raters grade different items. An analysis of variance of a crossed design yields estimates of seven variance components while a nested design yields five variance components. The variance components of the nested design can be obtained from the variance components of the crossed design. The reverse, deriving the variance components of a two-facet crossed design from a two-facet nested design, is not possible. For that reason, crossed designs are to be preferred over nested designs in a G-study. De Gruijter and Van der Kamp (1991, p. 55-56) make a case for the fact that a less informative design can turn out to be quite efficient.

Comparing equation (7) with equation (5), the contribution of $\sigma^2_{pj}$ is smaller for the nested than the crossed design. Thus, when the number of items and the number of raters per item is fixed, it is better to distribute the items among a larger group of raters than have only a few raters grade all the items. In educational research where the objects of measurement are often classes or schools, nesting is a 'natural' feature of the designs employed. A design in which persons are nested within classes and persons are crossed with items is commonly employed to answer the question of how well classes can be differentiated.

Another important modification concerns the sampling mode of the facets. If the conditions of a facet are assumed to be a random sample from an essentially infinite set of admissible conditions, the measures can be generalized to the entire set of conditions. However, assuming that the conditions of a facet constitute the entire set of admissible conditions limits the generalization of the measures to that fixed set of conditions. Restricting the universe of generalization by changing a random generalization facet into a fixed facet leads to a decrease of the error variance and an increase of the generalizability coefficient. Changing random facet $I$ of the two-facet random-model nested design into a fixed facet, for example, results in the generalizability coefficient for the two-facet mixed-model nested design being defined as

$$\mathcal{E}\rho^2 = \frac{\sigma^2_p + \frac{\sigma^2_{pi}}{n_i}}{\sigma^2_p + \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{pj,pij,e}}{n_i n_j}}. \qquad (8)$$

A comparison of generalizability coefficient (8) with generalizability coefficient (7) shows that the universe score variance of the two-facet random-model nested design is increased by adding the variance component for the person by fixed facet $I$ interaction divided by the

number of conditions of facet $I$. Generalizability coefficient (8) will therefore be at least equal to generalizability coefficient (7). It should be noted that while (8) is defined in terms of random-model components, it can also be defined in terms of mixed-model components.

Another type of modification, mentioned by Cardinet et al. (1981, p. 200), is to redefine the universe of generalization and/or the population of differentiation in order to reduce the contributions to error variance attributable to certain atypical conditions of a facet. After the elimination of atypical conditions, the generalization of the measures is of course restricted to a more restrained universe and/or population.

Changing a fixed facet into a random facet is justified only if the observed conditions can be considered interchangeable with other non-observed conditions of the same facet. To resolve the problem of whether a facet can be considered random or fixed, Shavelson and Webb (1981, p. 143) recommend applying the concept of *exchangeability*. This concept implies that even though conditions of a facet have not been sampled randomly, the facet may be considered to be random if conditions not observed can be exchanged with the observed conditions. Cardinet et al. (1981, p. 200) point out that the implications for the practical application of the measurement procedure should be taken into account before changing the sampling mode of a facet. For example, it is not realistic to assume that the same jury will be used in every administration of an examination.

While the usual objects of measurement in testing are persons, other facets of a measurement design can also be selected as objects of measurement. Cardinet and his associates (1981, 1983, 1985) use the terms *face of differentiation* to refer to facets corresponding to objects of measurement and *face of instrumentation* to refer to conditions of measurement. They also introduced the *principle of symmetry* to indicate that it is possible to transfer a facet of

differentiation to the face of instrumentation. For a design that crosses persons with items and crosses persons with raters this means that an investigator can ask not only how well persons can be differentiated, but can also ask how well raters can(not) be differentiated.

The most common design modification (Cardinet and Allal 1983, p. 42) is the one that increases the number of conditions of one or more instrumentation facets of the design in order to control the principal sources of random sampling error. In the following section this modification is discussed and illustrated for the designs described in section 1.2.

### 1.3.1 Optimization of the One-Facet Random-Model Crossed Design

To determine the minimum number of observations per person for the one-facet random model crossed design it suffices to rewrite (3) as

$$n = \frac{\rho^2 \sigma^2_{res}}{\sigma^2_p - \rho^2 \sigma^2_p}, \tag{9}$$

where $n$ is the number of observations per person needed to attain the stipulated value of $\rho^2$ (shorthand for $\mathcal{E}\rho^2$).
Equation (9) can be considered an alternative expression for the Spearman-Brown prophecy formula from classical test theory which states that an increase/decrease of the number of measures or observations results in an increase/decrease of the reliability coefficient.

As the Spearman-Brown prophecy formula is in essence an optimization problem, it can be stated in terms of mathematical programming as

| | | | |
|---|---|---|---|
| minimize | $L = n$ | objective-function | (10) |
| subject to | $\rho^2(n) = g$ | equality constraint. | (11) |

The common feature of mathematical programming models is that they all involve optimization, that is, the minimization or maximization of a quantity designated as the objective-function. In the minimization statement (10) of this optimization problem, $L$ refers to the value of the objective-function which results when different numbers of conditions, $n$, for facet $I$ are used. In constraint (11), $\rho^2(n)$ stands for the reliability coefficient of a one-facet random-model crossed design and $g$ for the value of a specific reliability coefficient.

Because of the equality constraint in (11), the value of the objective-function will often be non-integer. To account for the fact that the number of items is an integer by definition, the optimization problem is restated as

| | | | |
|---|---|---|---|
| minimize | $L = n$ | objective-function | (12) |
| subject to | $\rho^2(n) \geq g$ | threshold constraint | (13) |
| | $n$ integer | integer constraint. | (14) |

The introduction of threshold constraint (13) and integer constraint (14) will result in the minimum integer number of observations, for example, items or raters, needed to achieve a minimum acceptable reliability coefficient. Equations (12), (13), and (14) can be considered a generalization of the Spearman-Brown prophecy formula for one-facet designs.

### 1.3.2 Optimization of the Two-Facet Random-Model Crossed Design

In terms of mathematical optimization, the problem of determining the minimum number of observations for the two-facet random-model crossed design can be stated as

| | | | |
|---|---|---|---|
| minimize | $L = n_i n_j$ | objective-function | (15) |
| subject to | $\rho^2(n_i, n_j) \geq g$ | threshold constraint | (16) |
| | $n_i$ and $n_j$ integer | integer constraint. | (17) |

In the minimization statement (15) of this optimization problem, $L$ refers to the value of the objective-function which results when different numbers of conditions, $n_i$ and $n_j$, for facet $I$ and $J$ are used. In the threshold constraint (16), $\rho^2(n_i, n_j)$ stands for the generalizability coefficient of a two-facet random-model crossed design and $g$ for the minimum acceptable generalizability coefficient. The integer constraint (17) states that feasible values for $n_i$ and $n_j$ have to be integer values.

The determination of the minimum number of observations is much more complex for the two-facet random-model crossed design than for the one-facet random-model crossed design. The reason is that the error variance in generalizability coefficient (5) consists of three components, while the error variance in reliability coefficient (3) consists of only one component. A procedure for solving optimization problems for multi-facet random-model crossed designs is presented in Chapter 2.

The companion computer program referred to in Chapter 2 consists of a computer code for a branch-and-bound algorithm. It is important to note that the input of the program always assumes a crossed design. First, the number of facets and the threshold for the generalizability coefficient have to be specified. The variance components and an indication of the type of variance are the next entries. Three types of variance are distinguished: universe score variance, error variance, and variance which does not play a role in this particular optimization problem. The latter applies to $\sigma_i^2$, $\sigma_j^2$, and $\sigma_{ij}^2$, if the optimization problem relates to the generalizability coefficient defined by (5). The output of the program consists of the solutions which comply with the objective-function and constraints of the optimization problem and the corresponding search-tree. The computer program can be obtained from the author upon request.

### 1.3.3 Optimization of the Two-Facet Random-Model Nested Design

In terms of mathematical optimization, the formulation of the problem of determining the minimum number of observations for the two-facet random-model nested design does not differ from the formulation of the same problem for the two-facet random-model crossed design. What is different, however, is the definition of the threshold constraint in the two formulations. The error variance of the generalizability coefficient (5) of the random-model crossed design consists of three components, while the coefficient (7) for the random-model nested design consists of only two components. An example of a test with arithmetic items nested within three strata (25 items per stratum $h$), taken from Pilliner (1965) and analyzed by Cronbach et al. (1972, p. 226 ff.) will be used to illustrate the applicability of the computer program described before. The universe score variance, $\hat{\sigma}^2_p$, is equal to .0463. The variance component for the persons by strata interaction, $\hat{\sigma}^2_{ph}$ = .0030 plus the residual variance component, $\hat{\sigma}^2_{res} = \hat{\sigma}_{pi,pih,e}$ = .1572, constitute the two error variances. The number of facets is specified as two and the threshold for the generalizability coefficient is set at $g \geq .90$. Lower-bounds for the two facets are specified, $n_i \geq 1$ and $n_h \geq 1$. The universe score variance and the two error variances, $\hat{\sigma}^2_{ph}$ and $\hat{\sigma}^2_{pi,pih,e}$, are entered next. Since the computer program assumes a crossed design, the other variance components have to be indicated as not playing a role in this particular optimization problem. The optimal integer solutions for this optimization problem are $n_i$ = 7 items per stratum and $n_h$ = 5 strata.

## 1.4 Purpose of the Study

Classical test theory, restricted to one-facet designs with an undifferentiated error component, provides test constructors with the Spearman-Brown prophecy formula to determine the number of items or raters needed to obtain a specific reliability coefficient or to assess

the maximum reliability given a specific number of items. In generalizability theory, however, with its multi-faceted error component, 'the Spearman-Brown concept of reliability as a simple function of 'length of test' is not applicable to a multi-facet instrument' (Cronbach et al. 1972, p. 182). Unfortunately, Cronbach and his associates (1972) did not succeed in developing procedures to determine the best or optimal design for multi-facet measurement instruments. Woodward and Joe (1973) were the first to develop such procedures. Their procedures, however, have a number of shortcomings which will be discussed in the Chapters 2 and 3.

The current study was undertaken to provide test constructors with procedures to determine the optimal design for measurement instruments with more than one facet. To this end, procedures were developed for two pertinent test construction problems. In order to improve the economy of measurement instruments, the first problem addressed was the minimization of the number of observations given a minimum acceptable generalizability coefficient. As the quality of multi-facet measurement instruments depends upon the number of observations test constructors can make, the second problem addressed was the maximization of the generalizability coefficient of a measurement instrument given limited resources.

### 1.4.1 Overview of the Study

A procedure for determining the minimum number of observations per subject needed to achieve a specific generalizability coefficient is presented in Chapter 2. Optimal integer solutions for this optimization problem are obtained by means of a branch-and-bound algorithm that allows for the employment of constraints specified by the test constructor. The procedure is derived and illustrated for multi-facet random-model crossed designs, but, as has been shown in section 1.3.3,

can also be applied to designs with nested facets and mixed-model designs.

A procedure for maximizing the generalizability coefficient under the constraint of limited resources is presented in Chapter 3. Optimal integer solutions for this optimization problem are obtained in two steps. The solutions for a continuous relaxation of the constraint derived in the first step are used in the second step as the bounds in a branch-and-bound procedure to obtain the optimal integer solutions. The procedure is illustrated with the two-facet random-model crossed design.

In Chapter 4 another two-step procedure for maximizing the generalizability coefficient under a budget constraint is presented. This procedure has the advantage over the one in the foregoing chapter in that continuous solutions for designs with more than two facets can more easily be derived. The latter is accomplished by reducing a complex cost function to a 'simple' cost function that takes up a large amount of the budget. Solutions for this simple cost function appear to be excellent starting solutions for more complex cost functions. These starting solutions are used in the second step, consisting of a branch-and-bound procedure, to obtain the optimal integer solutions.

Another solution to the problem of maximizing the generalizability coefficient under a budget constraint is presented in Chapter 5. It is shown that for the two-facet random-model crossed design the Cauchy-Schwarz inequality can be applied to derive optimal continuous solutions which are an alternative to the solutions derived in Chapter 3. These continuous solutions can then again be used as the starting point for a branch-and-bound procedure to obtain the optimal integer solutions. It should be noted that the Cauchy-Schwarz inequality can also be applied to derive solutions for the minimization problem described in Chapter 2.

In Chapter 6 the problem of sampling variability of variance components in generalizability theory is addressed. The results of a study concerning this problem are examined.

In the Epilogue the main conclusions are formulated and topics for further research are indicated.

The material presented in chapters two to five consists of articles that have been published, accepted for publication or submitted for publication. As editorial notation requirements sometimes differ, the notation for the same equations and variables occassionally varies in the six chapters of this study.

## References


Adema, J.J. (1990). *Models and algorithms for the construction of achievement tests*. Enschede: FEBO.

Boekkooi-Timminga, E. (1989). *Models for computerized test construction*. De Lier: Academisch Boeken Centrum.

Brennan, R.L. (1983) *Elements of generalizability theory*. Iowa City: ACT.

Cardinet, J., & Allal, L. (1983). Estimation of generalizability parameters. In L.J. Fyans (Ed.), *Generalizability theory: Inferences and practical applications* (pp. 17-48). San Francisco: Jossey-Bass.

Cardinet, J., Tourneur, Y., & Allal, L. (1981). Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement, 18*, 183-204; *19*, 331-332.

Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure [Adequate Measurement]*. Bern: Peter Lang.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

de Gruijter, D.N.M., & van der Kamp, L.J.Th. (1991). Generalizability theory. In R.K. Hambleton, & J.N. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications* (pp. 45-68). Dordrecht: Kluwer.

Feldt, L.S., & Brennan, R.L. (1989). Reliability. In R.L. Linn (Ed.) *Educational Measurement* (3rd ed., pp. 105-146), Washington, DC: American Council on Education.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley. [Hillsdale: Erlbaum, 1987.]

Lord, F.M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement, 14*, 117-138.

Maxwell, A.E., & Pilliner, A.E.G. (1968). Deriving coefficients of reliability and agreement. *The British Journal of Mathematical and Statistical Psychology, 21,* 105-116.

Mellenbergh, G.J. (1977). The replicability of measures. *Psychological Bulletin, 84,* 378-384.

Pilliner, A.E.G. (1965). *The application of analysis of variance components in psychometric experimentation.* Unpublished doctoral dissertation. Edinburgh: University of Edinburgh.

Shavelson, R.J., & Webb, N.M. (1981). Generalizability theory: 1973-1980. *British Journal of Mathematical and Statistical Psychology, 34,* 133-166.

Theunissen, T.J.J.M. (1985). Binary programming and test design. *Psychometrika, 50,* 411-420.

Votaw, D.F. (1952). Methods of solving some personnel-classification problems. *Psychometrika, 17,* 255-266.

Woodward, J.A., & Joe, G.W. (1973). Maximizing the coefficient of generalizability in multi-facet decision studies. *Psychometrika, 38,* 173-181.

# Chapter 2

# Minimizing the Number of Observations: A Generalization of the Spearman-Brown Formula*

* The material in this chapter has been published in Sanders, P.F., Theunissen, T.J.J.M., and Baas, S.M. (1989). Minimizing the number of observations: A generalization of the Spearman-Brown formula. *Psychometrika, 54,* 587-598.

## 2.1 Introduction

In generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) a distinction is made between a generalizability (G) study and a decision (D) study. In a G-study, estimates of variance components are obtained which can be used by the investigator in a D-study. One of the major decisions an investigator has to make is how many observations per subject or another object of measurement are necessary in order to control the principal sources of random sampling error or to achieve a given generalizability coefficient (cf. Cardinet & Allal, 1983, p. 42).

For one-facet designs the minimum number of observations per subject can be determined as follows. The coefficient of reliability for the one-facet random-model crossed design, $\rho^2$, may be expressed as

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{res}^2}{n_1}}, \tag{1}$$

where $\sigma_p^2$ is the variance component for persons, $\sigma_{res}^2$ is the variance component for the $p$ x Facet 1 interaction plus the error, and $n_1$ is the number of observations, that is, conditions of Facet 1, in the D-study. Rewriting (1) and letting $\rho^2$ be a specific reliability coefficient, the minimum number of observations per subject is equal to

$$n_1 = \frac{\rho^2 \sigma_{res}^2}{\sigma_p^2 - \rho^2 \sigma_p^2}. \tag{2}$$

Equations (1) and (2) both exemplify the Spearman-Brown Prophecy Formula from classical test theory: increase/decrease of the number of observations, for example, items, results in an increase/decrease of the reliability coefficient. This correspondence between number of observations and reliability, however, does not extend to designs with more than one facet. Increasing the number of conditions of a facet with a large error variance component, for example, will have a greater impact on the generalizability coefficient than increasing the number of conditions of a facet with a small error variance component. With

multi-facet designs it is therefore possible to increase the generalizability coefficient while decreasing the number of observations. The multi-dimensional nature of error variance is the reason for this paradoxical result, which is inconsistent with assumptions in classical test theory but not with assumptions in generalizability theory (cf. Brennan, 1983, p. 67).

Because of the multi-dimensional nature of error variance in generalizability theory, the determination of the minimum number of observations is much more complex for multi-facet designs than for one-facet designs. Woodward and Joe (1973) presented a method for solving this problem, which they illustrated for the two-facet random-model crossed design. The generalizability coefficient for this design may be expressed as

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_p^2}{n_2} + \frac{\sigma_{res}^2}{n_1 n_2}}, \qquad (3)$$

where $\sigma_p^2$ is the variance component for persons, $\sigma_{p1}^2$ is the variance component for the person by Facet 1 interaction, $\sigma_{p2}^2$ is the variance component for the person by Facet 2 interaction, $\sigma_{res}^2$ is the variance component for the $p$ x Facet 1 x Facet 2 interaction plus error, $n_1$ and $n_2$ are the number of conditions of Facet 1 and Facet 2 in the D-study. Denoting the total number of observations for this design by $L = n_1 n_2$, the product of the number of conditions, (3) can be written as

$$L = \frac{\rho^2 (\sigma_{p1}^2 n_2 + \sigma_{res}^2)}{\sigma_p^2 - \rho^2 \sigma_p^2 - \frac{\rho^2 \sigma_{p2}^2}{n_2}}. \qquad (4)$$

Woodward and Joe's method consists of taking the derivative of this function with respect to $n_2$ and setting the result equal to zero, thus obtaining the following quadratic equation

$$0 = (\sigma_p^2\sigma_{p1}^2)\,\frac{1-\rho^2}{\rho^2}\,n_2^2 - (2\sigma_{p1}^2\sigma_{p2}^2)\,n_2 - \sigma_{res}^2\sigma_{p2}^2. \qquad (5)^*$$

The positive root of the equation is taken as the desired solution and $L$ is found by substituting the positive $n_2$ into (4). The results of their method for a two-facet random-model crossed design with $\hat{\sigma}_p^2 = 5.435$, $\hat{\sigma}_{p1}^2 = 3.421$, $\hat{\sigma}_{p2}^2 = 1.140$ and $\hat{\sigma}_{res}^2 = 11.850$ are presented in Table 1.

**TABLE 1**

Values of Minimum $L$ and Optimal Values of $n_1$ and $n_2$ for Various Values of $\rho^2$

| $\rho^2$ | $L$ | $n_1$ | $n_2$ |
|---|---|---|---|
| .97 | 685.9 | 45.4 | 15.1 |
| .96 | 402.1 | 34.7 | 11.6 |
| .95 | 267.1 | 28.7 | 9.4 |
| .94 | 192.0 | 24.0 | 8.0 |
| .93 | 145.4 | 20.9 | 6.9 |
| .92 | 114.5 | 18.5 | 6.2 |
| .91 | 92.8 | 16.7 | 5.6 |
| .90 | 77.0 | 15.2 | 5.1 |
| .89 | 65.1 | 13.9 | 4.6 |
| .88 | 55.8 | 12.9 | 4.3 |
| .87 | 48.4 | 12.1 | 4.0 |
| .86 | 42.5 | 11.3 | 3.8 |
| .85 | 37.6 | 10.6 | 3.5 |
| .84 | 33.5 | 10.0 | 3.3 |
| .83 | 30.1 | 9.5 | 3.2 |
| .82 | 27.0 | 9.0 | 3.0 |
| .81 | 24.7 | 8.6 | 2.8 |
| .80 | 22.5 | 8.2 | 2.7 |
| .79 | 20.6 | 7.8 | 2.6 |
| .78 | 18.9 | 7.5 | 2.5 |
| .77 | 17.5 | 7.2 | 2.4 |
| .76 | 16.1 | 6.9 | 2.3 |
| .75 | 14.9 | 6.7 | 2.2 |
| .74 | 13.9 | 6.4 | 2.2 |
| .73 | 13.0 | 6.2 | 2.1 |
| .72 | 12.1 | 6.0 | 2.0 |
| .71 | 11.3 | 5.8 | 1.9 |
| .70 | 10.6 | 5.6 | 1.8 |

* In Woodward and Joe (Psychometrika, 1973, p. 177) and Sanders et al. (Psychometrika, 1989, p. 588) this equation contains errors.

The method proposed by Woodward and Joe, however, has a number of shortcomings. Firstly, the method requires derivations and calculations which can become quite complex for designs with more than two facets, nested designs and designs based on other models. Secondly, the values obtained for the number of conditions of different facets and $L$ are usually non-integer. Woodward and Joe's solution for this problem consists of rounding off the values obtained for $n_1$ and $n_2$ to the nearest whole numbers. By rounding off, however, their conclusion that the positive root of the quadratic equation will always give the smallest $L$ no longer applies. Consequently, the values for $n_1$ and $n_2$ resulting from their method will not always be the optimal values. For instance, specifying a value of .79 for the generalizability coefficient and rounding off the values of $n_1$ and $n_2$ to 8 and 3 will result in $L$ being equal to 24. The specified value of .79, however, could also have been obtained with 7 conditions for Facet 1 and 3 conditions for Facet 2, resulting in $L = 21$ observations. (The generalizability coefficients for these and other values for the number of conditions referred to in this article are presented in Table 2.)

**TABLE 2**

Various Values of $n_1$, $n_2$, $L$ With Resulting Variance Components and $\rho^2$

| $n_1$ | $n_2$ | $L$ | $\hat{\sigma}^2_p$ | $\frac{\hat{\sigma}^2_{p1}}{n_1}$ | $\frac{\hat{\sigma}^2_{p2}}{n_2}$ | $\frac{\hat{\sigma}^2_{res}}{n_1 n_2}$ | $\rho^2$ |
|---|---|---|---|---|---|---|---|
| 6 | 4 | 24 | 5.4 | .57017 | .285 | .49375 | .80166 |
| 6 | 6 | 36 | 5.4 | .57017 | .190 | .32917 | .83303 |
| 7 | 3 | 21 | 5.4 | .48871 | .380 | .56429 | .79135 |
| 7 | 4 | 28 | 5.4 | .48871 | .285 | .42321 | .81952 |
| 8 | 3 | 24 | 5.4 | .42763 | .380 | .49375 | .80681 |
| 9 | 3 | 27 | 5.4 | .38011 | .380 | .43889 | .81926 |
| 10 | 3 | 30 | 5.4 | .34210 | .380 | .39500 | .82951 |
| 11 | 3 | 33 | 5.4 | .31100 | .380 | .35909 | .83808 |
| 12 | 2 | 24 | 5.4 | .28508 | .570 | .49375 | .80117 |
| 24 | 2 | 48 | 5.4 | .14254 | .570 | .24688 | .84996 |
| 36 | 2 | 72 | 5.4 | .09503 | .570 | .16458 | .86757 |

A difference of one condition for one facet can make a substantial difference in resources. For instance, with a two-facet crossed design the difference of one condition could mean that one rater less is needed to correct the answers of a hundred students to ten questions. That the generalizability coefficients of the two designs hardly differ is due to the insensitivity of higher values of the coefficient to even major changes in the design. It is clear that for designs with more than two facets, the difference of one condition for one facet can make an even more dramatic difference in resources. A third and more serious shortcoming of the method, however, is that it implies a very restrictive constraint on the values for $n_1$ and $n_2$. Table 1 shows that the values for $n_1$ and $n_2$ are determined by the ratio of the variance components $\sigma^2_{p1}$ and $\sigma^2_{p2}$, that is, the ratio $\hat{\sigma}^2_{p1} / \hat{\sigma}^2_{p2} = 3.421/1.140$. Therefore more than three times as many conditions are allocated to $n_1$ than to $n_2$.

The method proposed here is different from Woodward and Joe's method in a practical as well as theoretical respect. The new method consists of an algorithm based on the concept of enumeration, using a branch-and-bound algorithm, the principles of which are well known in integer programming (e.g., Salkin, 1975). The method is presented in three parts. First, the structure of a branch-and-bound algorithm is described. Next, an algorithm for a two-facet random-model crossed design is presented. Finally, the algorithm for the two-facet design is generalized for multi-facet designs.

## 2.2 Branch-and-Bound Algorithm

The term branch-and-bound algorithm does not refer to one specific algorithm but to a class of algorithms. Papadimitriou and Steiglitz (1982, p. 433) describe the branch-and-bound approach as the construction of a proof that a solution is optimal based on successive partitioning of the solution space. The parts branch and bound refer to

rules which reduce the amount of search to be conducted for the optimal solution. A branch-and-bound algorithm is usually represented by a tree composed of branches and nodes, with the nodes organized in levels. In the tree as it is organized for the problem considered here, level $\ell, \ell = 1,2,\ldots,t$, corresponds with variable $n_\ell$. Each node at Level $\ell$ represents a partial solution in which variables $n_1, n_2, \ldots, n_\ell$ have fixed values, say $n_i = \hat{n}_i, i = 1,2,\ldots,\ell$, whereas the remaining $t - \ell$ variables are said to be free, to indicate that their values still need to be determined in the further course of the search-process. The root of the full search-tree consists of a single node at Level 0 in which all variables are free. The nodes at the highest level (level $t$) correspond with complete solutions of the problem and may therefore all be regarded as candidates for the optimal solution. A node at level $\ell \leq t - 1$, with partial solution $\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_\ell$, is connected by branches with all nodes at level $\ell + 1$ in which the associated partial solution is different only in the fixed value of variable $n_{\ell+1}$, which was free before. Starting from the root of the tree, a complete solution can gradually be developed by passing through individual nodes, one at each level, until level $t$ is reached. In this way any specific node at level $t$ is reachable along a unique path from the root. On the other hand, from any node at an intermediate level $\ell < t$ several nodes at level $t$ may be attained. In general the tree and thus the number of complete solutions will increase rapidly as the number of variables increases. What is needed therefore are additional rules that will allow a significant reduction of the full search-tree by cutting off those parts which are irrelevant regarding an optimal solution.

## 2.3 Branch-and-Bound for the Two-Facet Design Problem

The problem for the two-facet design can be stated in terms of mathematical optimization as

| | | | |
|---|---|---|---|
| minimize | $L = n_1 n_2$ | objective-function | (6) |
| subject to | $\rho^2(n_1, n_2) \geq g$ | threshold constraint | (7) |
| | $n_1 \geq n_2 \geq 1$ | monotonicity constraints | (8) |
| | $n_1$ and $n_2$ integer | integer constraint. | (9) |

In the minimization statement (6) of this optimization problem, $L$ refers to the value of the objective-function which results when different numbers of conditions, $n_1$ and $n_2$, for Facet 1 and 2 are used. Note that $L$ can also be interpreted as minimizing the cost of $L$ observations. If $c_1$ and $c_2$ are the costs associated with $n_1$ and $n_2$ respectively, then $L = c_1 c_2 n_1 n_2$ corresponds with the total cost of $L$ observations. However, since $c_1 c_2$ is a constant, it will not influence the result of the minimization procedure and may therefore be omitted.

In the threshold constraint (7), $\rho^2(n_1, n_2)$ stands for the generalizability coefficient of a two-facet random-model crossed design and $g$ for the lowest acceptable value of a generalizability coefficient. The function $\rho^2(n_1, n_2)$ is strictly increasing with respect to both variables.

The monotonicity constraint (8) $n_1 \geq n_2$ employed here is but one of many linear inequality constraints that could be employed. Note that an optimal solution for the two-facet design problem can also be obtained without this constraint. However, an algorithm employing this constraint will exclude an irrelevant part of the decision-space and consequently reduce the number of branchings in the branch-and-bound process described hereafter.

The integer constraint (9) states that feasible values for $n_1$ and $n_2$ have to be integer values.

After the problem has been formulated as an optimization problem,

bounding rules are constructed which effectively reduce the search-process. A distinction can be made between feasibility bounds, that is, bounds on the values that $n_1$ and $n_2$ can assume without violating the constraints, and optimality bounds, bounds that use a comparison of objective-function values to ascertain whether a given partial solution can lead to an optimal solution. Both types of bounds will be derived in this paper.

To fix the number of relevant branches emanating from the root of the search-tree, a lower bound $\ell b_1$ and an upper bound $ub_1$ on the values that $n_1$ can assume in an optimal solution can be derived as follows. Regarding the threshold constraint, it can easily be seen that $\rho^2(n_1, n_2)$ is strictly increasing both in $n_1$ and $n_2$ so that if $n_1 \geq n_1^*$ and $n_2 \geq n_2^*$, then $\rho^2(n_1, n_2) \geq \rho^2(n_1^*, n_2^*)$ while strict inequality holds whenever $n_1 > n_1^*$ or $n_2 > n_2^*$. Hence for a given value $\hat{n}_1$ of $n_1$, there either exists a least integer value $\hat{n}_2 \leq \hat{n}_1$ that the threshold constraint is satisfied, or for all $n_2 \leq \hat{n}_1$ this constraint is violated. This observation implies the existence of a value $n^-$ such that $\rho^2(n_1 = n^-, n_2 = n^-) \geq g$, whereas for all $n_1 < n^-$ no value $n_2 \leq n_1$ exists that satisfies the threshold constraint. This means that $n^-$ should be taken as the lower bound for Facet 1. For instance, with (3) it can be calculated that for Woodward and Joe's example $n_1 = 6$ is the lowest integer value such that $\rho^2$ (6,6) $\geq .80$, while for $n_1 \leq 5$ and for each value $n_2 \leq n_1$ one finds $\rho^2(n_1, n_2) < .80$. Consequently, in this case $\ell b_1 = n^- = 6$. Note that without the integer constraint $n_1 = n_2 = n^- = 5.08$.

Let $L^- = n_1^- n_2^-$ be the value of the objective-function associated with the solution $n_1 = n_2 = n^-$. Then $ub_1 = L^-$ is an obvious upper bound on the values of $n_1$, because any value $n_1 > ub_1$ produces a solution that has a value $n_1 n_2 > ub_1 = L^-$ and is therefore irrelevant for optimality. While initially $ub_1$ may thus be intractably high, its value can be adapted throughout the search-process whenever a new and better solution is found. The current best solution is called the

incumbent. So $n_1 = n_2 = n^-$ as defined above is the initial incumbent. As soon as a complete solution $(n_1^*, n_2^*)$ with value $L^* = n_1^*, n_2^* < L^-$ is derived, this becomes the new incumbent and $ub_1 = L^*$ replaces the initial upper bound. This process with repeated successive adaptations of both the incumbent and $ub_1$ is continued until the whole search-tree has been explored. The then operative incumbent is designated as the optimal solution.

In the example, $ub_1 = n^- n^- = 36$ becomes the first upper bound for $n_1$. With $\ell b_1 = 6$ and $ub_1 = 36$ the total number of nodes at Level 1 becomes $ub_1 - \ell b_1 + 1 = 31$. Any such node represents a partial solution with a fixed value for $n_1$ and $n_2$ free (see Figure 1). Now, for any node at Level 1 corresponding with a fixed value $\hat{n}_1$ for $n_1$ there is a least integer value $\hat{n}_2$ for $n_2$ such that $\rho^2(\hat{n}_1, \hat{n}_2) \geq g$. Since this value minimizes $L = \hat{n}_1 n_2$ subject to the constraints $\rho^2(\hat{n}_1, n_2) \geq g$ and $\hat{n}_1 > n_2 \geq 1$, it suffices to consider a tree in which each node at Level 1 corresponds with precisely one node at Level 2. For instance, with (3) the unique feasible value for $n_2$ that minimizes $\hat{n}_1 n_2$ given $\hat{n}_1 = 8$ is found to be $\hat{n}_2 = 3$.

By convention, the branch associated with the lower bound is further referred to as 'left-most' branch and the one associated with the upper bound as 'right-most' branch. All branches between Level 0 (the root) and Level 1 are further ordered according to increasing values of $n_1$. The search-tree for the two-facet problem is now traversed as follows. Starting at the root of the tree, the left-most branch to Level 1 is considered first, immediately followed by the unique branch that connects the node with partial solution $\hat{n}_1 = \ell b_1$ to the corresponding node at Level 2 with complete solution $n_1 = \hat{n}_1 = \ell b_1, n_2 = \hat{n}_2$. Next the process returns to the root, which for the two-facet problem is the most recent node from which not all branches to the next level have been considered yet. This is represented by saying that the root is not fathomed. Note that each

node at Level 1 is fathomed as soon as the search-process has passed through it once, due to the fact that there is only one branch leading to Level 2.

The two nodes regarded next are the one at Level 1 with partial solution $\hat{n}_1 = \ell b_1 + 1$, and its corresponding node at Level 2. The full search-process as it is conducted for the example with $g \geq .80$ is shown in Figure 1. In this figure the nodes are numbered in their order of appearance.

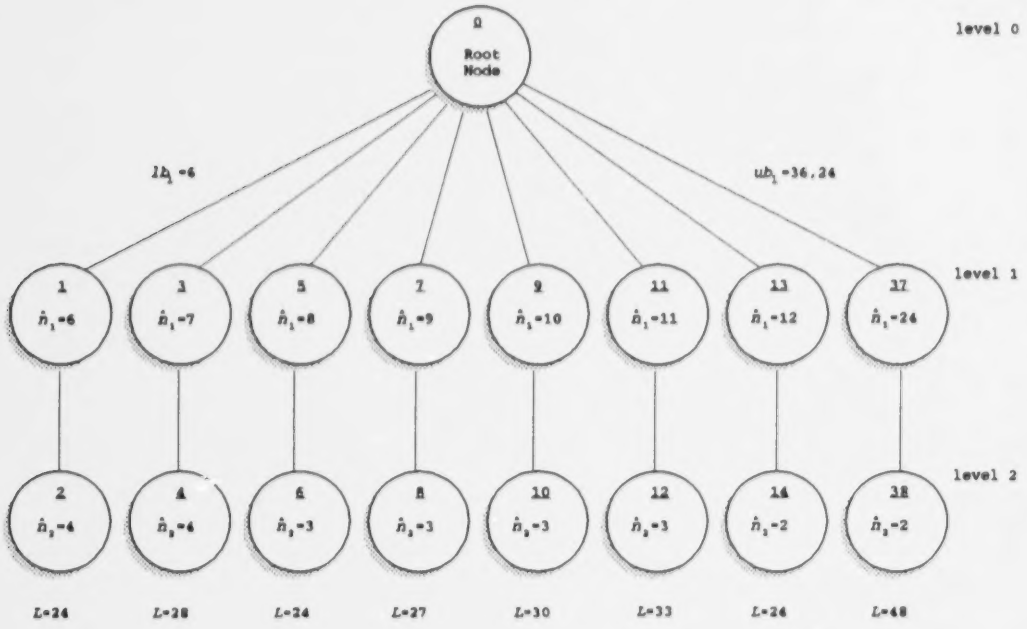


**FIGURE 1**

Search-tree for the two-facet example

The initial incumbent, whose value is 36, is replaced in Node 2 by $(\hat{n}_1, \hat{n}_2) = (6,4)$, giving the new incumbent the value 24. Further exploration of the search-tree renders two more solutions, producing the same value for the objective-function as Node 2: $(\hat{n}_1, \hat{n}_2) = (8,3)$ in Node 6 and $(\hat{n}_1, \hat{n}_2) = (12,2)$ in Node 14. The search-process ends in Node 38 with solution $(\hat{n}_1, \hat{n}_2) = (24,2)$, yielding the objective-function value 48 which is higher than the incumbent.

A closer inspection of the search-tree shown in Figure 1 reveals that a considerable reduction can be obtained regarding the size of the tree by eliminating irrelevant nodes and branches. For example, all nodes from 15 up to and including 38 are superfluous, as are all corresponding branches between Level 1 and 2, since they give rise to solutions in which $n_2$ remains constant at a value $\hat{n}_2 = 2$, implying objective-function values $\geq 13$ times $2 = 26 > 24$. This observation is clearly related to the fact that from each node at Level 1 there is a unique branch passing to Level 2. This allows for the formulation of the following optimality bound. Let the values of $n_1$ at Level 1, further referred to as $\hat{n}_{1,i}, i = 1,2,\ldots,s = ub_1 - \ell b_1 + 1$, be such that $\hat{n}_{1,1} = \ell b_1, \hat{n}_{1,2} = \ell b_1 + 1, \ldots, \hat{n}_{1,i} = \ell b_1 + i - 1, \ldots, \hat{n}_{1,s} = ub_1$. Accordingly, for $i = 1,2,\ldots,s$, $\hat{n}_{2,i}$ is the value of $n_2$ that is uniquely associated with the partial solution $\hat{n}_{1,i}$, in a way described before. Then as long as $\hat{n}_{2,i}$ remains unaltered while $\hat{n}_{1,i}$ steadily increases by values of $i$, there will certainly be no improvement in the value of the objective-function. The objective-function value can only be improved when $\hat{n}_{2,i} > \hat{n}_{2,i+1}$ for some $i$. So let $i_0, i_1, \ldots, i_q$ be such that $i_0 = 1$ and $\hat{n}_{2,i_0} = \hat{n}_{2,1} = \ldots = \hat{n}_{2,i_1-1} > \hat{n}_{2,i_1} = \hat{n}_{2,i_1+1} = \ldots = \hat{n}_{2,i_2+1} = \ldots = \hat{n}_{2,i_2-1} > \hat{n}_{2,i_2} = \ldots = \hat{n}_{2,i_q-1} > \hat{n}_{2,i_q} = \hat{n}_{2,i_q+1} = \ldots = \hat{n}_{2,s}$. For our example it can be inferred that $4 = \hat{n}_{2,1} = \hat{n}_{2,2} > 3 = \hat{n}_{2,3} = \hat{n}_{2,4} = \hat{n}_{2,5} = \hat{n}_{2,6} > 2 = \hat{n}_{2,7} = \ldots = \hat{n}_{2,19}$. The lowest index $i$ for which $\hat{n}_{2,i} = 3$ is therefore equal to $i_1 = 3$ and the lowest index for which $\hat{n}_{2,i} = 2 = \hat{n}_{2,s}$ is equal to $i_2 = 7$, implying $q = 2$. These results give rise to a search-tree in which only the nodes at Level 1 corresponding with values $\hat{n}_{1,i_0}, \hat{n}_{1,i_1}, \ldots, \hat{n}_{1,i_q}$ and the associated nodes at Level 2 are considered. The values of $i_0, i_1, \ldots, i_q$ can be derived as follows. After $\hat{n}_{1,1} = \ell b_1$ has been established, the associated value of $\hat{n}_{2,1}$ as the least integer value such that $\rho^2(\hat{n}_{1,1}, \hat{n}_{2,1}) \geq g$ can be calculated with (3). Let $ub_2 = \hat{n}_{2,1}$. Now regarding a value $ub_2 - 1$ for $n_2$, $i_1$ is chosen as the lowest index such that $\hat{n}_{1,i_1} \leq ub_1 = \hat{n}^2_{1,1}$ and

$\rho^2(\hat{n}_{1,i_1}, ub_2 - 1) \leq g$. By generalizing this rule, the value of $i_j$ becomes the lowest index such that $\hat{n}_{1,i_j} \leq ub_1$ and $\rho^2(\hat{n}_{1,i_j}, ub_2 - j) \geq g$, $j = 0,1,\ldots,ub_2 - q(q \geq 1)$. Using this optimalitity bound in our example results in a reduced search-tree which is presented in Figure 2, where the numbering of the nodes refers to the nodes in Figure 1.

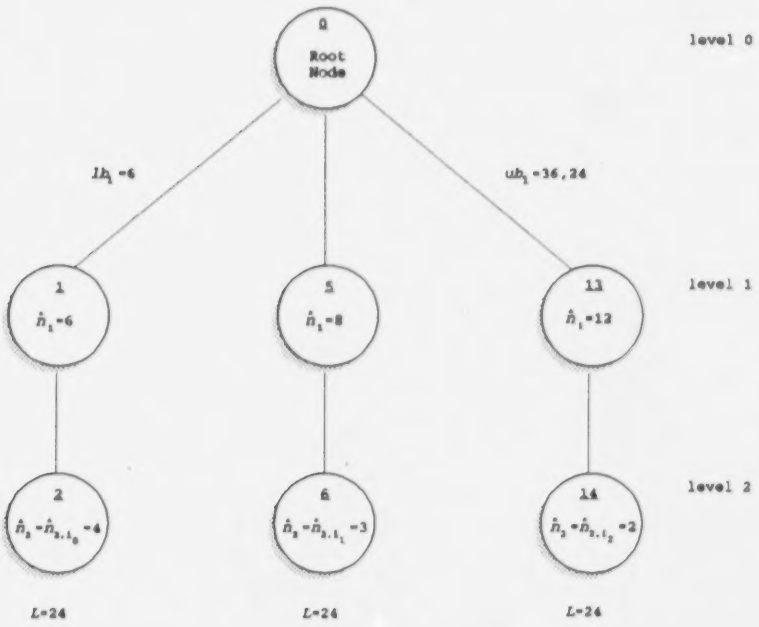


**FIGURE 2**

Reduced search-tree for the two-facet example ($i_0$=1, $i_1$=3, $i_2$=7)

At the end of the search-process there appear to be three candidates for an optimal solution: $(\hat{n}_1, \hat{n}_2)$ = (6,4) , $(\hat{n}_1, \hat{n}_2)$ = (8,3) and $(\hat{n}_1, \hat{n}_2)$ = (12,2). Solution $(\hat{n}_1, \hat{n}_2)$ = (8,3) could be considered the 'most optimal' solution because it results in a higher generalizability coefficient than the other two solutions. However, considerations other than obtaining a specific generalizability coefficient can and often will play a role when constructing a measurement instrument. If in this example Facet 1 had been items and Facet 2 raters, there could have

been considerable differences in the costs per condition of these two facets. Raters probably being more expensive than items, an investigator could for economic reasons prefer to increase the number of items rather than to increase the number of raters. This should be indicated by employing an economic constraint such as $n_1 \geq 5n_2$ instead of constraint $n_1 \geq n_2$ which is a psychometric constraint. Using the above specification $g \geq .80$ and employing the economic constraint mentioned, the optimal solution appears to be $(\hat{n}_1, \hat{n}_2) = (12,2)$.

### 2.4 Branch-and-Bound for the Multi-Facet Design Problem

The branch-and-bound method developed for the two-facet problem can be generalized for the multi-facet problem as follows

| | | |
|---|---|---|
| minimize | $L(n_1, n_2, \ldots n_t) =$ | |
| | $\prod_{i=1}^{t} n_i = n_1 n_2 \ldots n_t$ | objective function |
| subject to | $\rho^2(n_1, n_2, \ldots, n_t) \geq g$ | threshold constraint |
| | $n_1 \geq n_2 \geq \ldots \geq n_t \geq 1$ | monotonicity constraints |
| | $n_1, n_2, \ldots, n_t$ integer | integer constraint. |

The level-wise organization of a search-tree for a branch-and-bound algorithm to solve this multi-facet problem is analogous to the organization of the search-tree for the two-facet problem in the foregoing section. Let $k$ be an arbitrary node at level $\ell$, with associated partial solution $\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_\ell$. The branches passing from this node to level $\ell + 1$ represent feasible values for $n_{\ell+1}$, given the partial solution. The most important question to be answered next is how to obtain both a lower bound $lb(\ell, k)$ and an upper bound $ub(\ell, k)$ on relevant values for $n_{\ell+1}$. Writing $\rho^2_\ell(n_{\ell+1}, n_{\ell+2}, \ldots, n_t) = \rho^2(\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_\ell, n_{\ell+1}, n_{\ell+2}, \ldots, n_t)$ let $n^-$ be the least integer value such that $\rho^2_\ell(n^-, n^-, \ldots, n^-) \geq g$. Then, using the fact that $\rho^2_\ell$ is strictly increasing in all its arguments, it can be easily verified

that for each $n_{\ell+1} < n^-$ there exists no feasible completion of the partial solution $\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_\ell$. For if such a completion existed, with $n_i = \hat{n}_i, i = \ell + 1, \ell + 2, \ldots t$ and $n^- > \hat{n}_{\ell+1} \geq \hat{n}_{\ell+2} \geq \ldots \geq \hat{n}_t \geq 1$, then a 'constant' completion with $n_i = \hat{n}_{\ell+1}, i = \ell+1, \ell+2, \ldots, t$ would also be feasible, especially since $\rho_i^2(\hat{n}_{\ell+1}, \hat{n}_{\ell+1}, \ldots, \hat{n}_{\ell+1}) \geq \rho_i^2(\hat{n}_{\ell+1}, \hat{n}_{\ell+2}, \ldots, \hat{n}_t) \geq g$. This obviously contradicts the choice of $n^-$ as a least constant integer value yielding a feasible completion. Consequently, $lb(\ell, k) = n^-$ can be regarded as a proper lower bound. In conformity with the two-facet design problem, an initial upper bound $ub(\ell, k)$ is obtained by putting $n_{\ell+1} = n_{\ell+2} = \ldots = n_t = n^-$ yielding $ub(\ell, k) = \hat{n}_1 \hat{n}_2 \hat{n}_3 \ldots \hat{n}_\ell n^- \ldots n^- n^-$ ($= t - \ell$ times $n^-$). This upper bound is adapted each time a lower objective-function value is obtained.

To realize a further reduction in the size of the tree consider a node $k$ at level $t-2$. This node may be regarded as the root of a sub-tree in which the nodes at level $t-1$ correspond with values satisfying $lb(t-2, k) \leq n_t \leq ub(t-2, k)$*. In fact, a situation occurs that does not differ essentially from the two-facet case in the previous section. Using arguments similar to those used in that case, it is possible to determine for each node at level $t-1$ belonging to the sub-tree, a unique node at level $t$ that represents an optimal completion of the associated partial solution. Moreover, the number of relevant nodes in the sub-tree at level $t-1$ and thus the number of branches between level $t-2$ and level $t-1$ may again be considerably reduced using a type of optimality bound similar to that used for the two-facet case.

The foregoing considerations give rise to straightforward generalizations of the principles underlying the branch-and-bound algorithm for the two-facet problem. They actually fix the structure of the search-tree for the multi-facet problem. What remains to be

---

* In Sanders et al. (Psychometrika, 1989, p. 595) this sentence contains an error.

elucidated is how to perform subsequent steps such that in the end the full tree is most efficiently traversed. For that purpose a so-called depth-first strategy appears most plausible. This principle can be simply interpreted as follows. Let node $k$ at level $\ell < t$ be the node that the branching process has just reached. Then, before the process returns to the node at level $\ell-1$ immediately preceding $k$, the full sub-tree rooted in $k$ is explored. If, in addition, the order in which the branches are covered is agreed upon, for example from 'left' (the branch associated with $\ell b(\ell,k)$) to 'right' ($ub(\ell,k)$), the course of the search-process is completely fixed. As in the two-facet case an incumbent, or current best solution, is retained and adapted whenever relevant. To initialize a constant solution $n_i = n^-, i = 1,2,\ldots t$, with $n^-$ the least integer value such that $\rho^2(n^-,n^-,\ldots,n^-) \geq g$, suggests itself, in which $n^-$ corresponds with the lower bound on values of $n_1$.

The branch-and-bound method for the multi-facet problem is illustrated by the three-facet example described in Cronbach (1972, p. 171 ff.), where the aphasic symptoms of 30 patients are rated by 4 raters (Facet 2) on 6 graphic subtests (Facet 3) using the same 10 objects (Facet 1).

The generalizability coefficient for the three-facet random-model crossed design may be expressed as

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2} + \frac{\sigma_{p3}^3}{n_3} + \frac{\sigma_{p12}^2}{n_1 n_2} + \frac{\sigma_{p13}^2}{n_1 n_3} + \frac{\sigma_{p23}^2}{n_2 n_3} + \frac{\sigma_{res}^2}{n_1 n_2 n_3}}, \qquad (10)$$

where $\sigma_p^2$ is the variance component for patients and the other components are the interaction components divided by the number of conditions being used in the D-study.

The problem for the three-facet design is stated as

| | | | |
|---|---|---|---|
| minimize | $L = n_1 n_2 n_3$ | objective-function | (11) |
| subject to | $\rho^2(n_1 n_2 n_3) \geq g$ | threshold constraint | (12) |
| | $n_3 \geq n_1 \geq n_2 \geq 1$ | monotonicity constraints | (13) |
| | $n_1, n_2$ and $n_3$ integer | integer constraint. | (14) |

Note that the remarks made on the monotonicity constraints employed in the two-facet example above also apply for the monotonicity constraints employed in the three-facet example.

The results for this example with $\hat{\sigma}_p^2 = 5.3$, $\hat{\sigma}_{p1}^2 = .41$, $\hat{\sigma}_{p2}^2 = .15$, $\hat{\sigma}_{p3}^2 = 1.71$, $\hat{\sigma}_{p12}^2 = .15$, $\hat{\sigma}_{p13}^2 = 2.01$, $\hat{\sigma}_{p23}^2 = .07$ and $\hat{\sigma}_{res}^2 = .99$ and $g \geq .90$ are presented in Table 3.

**TABLE 3**

Three Different Constraints With Resulting Values for $n_1$, $n_2$, $n_3$, $L$, Variance Components and $\rho^2$

| $n_1$ | $n_2$ | $n_3$ | $L$ | $\hat{\sigma}_p^2$ | $\frac{\hat{\sigma}_{p1}^2}{n_1}$ | $\frac{\hat{\sigma}_{p2}^2}{n_2}$ | $\frac{\hat{\sigma}_{p3}^2}{n_3}$ | $\frac{\hat{\sigma}_{p12}^2}{n_1 n_2}$ | $\frac{\hat{\sigma}_{p13}^2}{n_1 n_3}$ | $\frac{\hat{\sigma}_{p23}^2}{n_2 n_3}$ | $\frac{\hat{\sigma}_{p123}^2}{n_1 n_2 n_3}$ | $\rho^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_3 \geq n_1 \geq n_2 \geq 1$ | | | | | | | | | | | | |
| 5 | 3 | 5 | 75 | 5.3 | .08200 | .05000 | .34200 | .01000 | .08040 | .00467 | .01320 | .90101 |
| 4 | 3 | 6 | 72 | 5.3 | .10250 | .05000 | .28500 | .01250 | .08375 | .00389 | .01375 | .90577 |
| 5 | 2 | 6 | 60 | 5.3 | .08200 | .07500 | .28500 | .01500 | .06700 | .00583 | .01650 | .90655 |
| 4 | 2 | 7 | 56 | 5.3 | .10250 | .07500 | .24429 | .01875 | .07179 | .00500 | .01768 | .90831 |
| 6 | 1 | 7 | 42 | 5.3 | .06800 | .15000 | .24429 | .02500 | .04786 | .01000 | .02357 | .90301 |
| 5 | 1 | 8 | 40 | 5.3 | .08200 | .15000 | .21375 | .03000 | .05025 | .00875 | .02475 | .90451 |
| 4 | 1 | 9 | 36 | 5.3 | .10250 | .15000 | .19000 | .03750 | .05583 | .00778 | .02750 | .90273 |
| 3 | 1 | 12 | 36 | 5.3 | .13666 | .15000 | .14250 | .05000 | .05583 | .00583 | .02750 | .90315 |
| $n_1 \geq n_3 \geq n_2 \geq 1$ | | | | | | | | | | | | |
| 5 | 3 | 5 | 75 | 5.3 | .08200 | .05000 | .34200 | .01000 | .08040 | .00467 | .01320 | .90101 |
| 6 | 2 | 5 | 60 | 5.3 | .06800 | .07500 | .34200 | .01250 | .06700 | .00700 | .01650 | .90014 |
| 7 | 1 | 7 | 49 | 5.3 | .05857 | .15000 | .24429 | .02143 | .04102 | .01000 | .02020 | .90668 |
| 8 | 1 | 6 | 48 | 5.3 | .05125 | .15000 | .28500 | .01875 | .04187 | .01167 | .02062 | .90149 |
| $n_1 \geq n_2 \geq n_3 \geq 1$ | | | | | | | | | | | | |
| 5 | 5 | 5 | 125 | 5.3 | .08200 | .03000 | .34200 | .00600 | .08040 | .00280 | .00792 | .90581 |

To illustrate our method, three different constraints were employed. The solutions in Table 3 show that the psychometric

constraint $n_3 \geq n_1 \geq n_2 \geq 1$ results in $L = 36$ observations. This number of observations corresponds with solutions $(\hat{n}_1, \hat{n}_2, \hat{n}_3) = (4,1,9)$ and $(\hat{n}_1, \hat{n}_2, \hat{n}_3) = (3,1,12)$.

However, these solutions are also the most expensive solutions since they involve the construction of new tests. An investigator who will therefore probably want to use the six tests that are already available should add constraint $n_3 \leq 6$ to this constraint, obtaining as an optimal solution $(\hat{n}_1, \hat{n}_2, \hat{n}_3) = (5,2,6)$ with $L = 60$. A more economical instrument would be one composed of as many objects and as few raters as possible and using no more tests than are available. Employing the corresponding economic constraints $n_1 \geq n_3 \geq n_2 \geq 1$ and $n_3 \leq 6$ would result in the optimal solution $(\hat{n}_1, \hat{n}_2, \hat{n}_3) = (8,1,6)$ with $L = 48$. Employing another economic constraint $n_1 \geq n_2 \geq n_3 \geq 1$ would result in the optimal solution $(\hat{n}_1, \hat{n}_2, \hat{n}_3) = (5,5,5)$ with $L = 125$ observations.

## 2.5 Conclusions and Discussion

The method proposed in this article enables an investigator to specify an acceptable threshold for generalizability coefficients. The employment of the threshold constraint together with the integer constraint necessarily results in values for the objective-function and values for the number of conditions of facets that are integer. Woodward and Joe's method requires the investigator to specify an exact value for the generalizability coefficient which will almost always lead to non-integer solutions for the number of conditions of the facets. However, having an investigator specify a threshold for the generalizability coefficient, which is done in our method, is much more realistic. It is more likely that an investigator can specify a minimum value for a generalizability coefficient than a specific value for a generalizability coefficient.

The method proposed here employs ordinal, threshold and monotonicity constraints. Employing these constraints results in

obtaining all the integer solutions which are in accord with the specified objective-function value and can be considered optimal solutions. On the other hand, Woodward and Joe's method implies a restrictive ratio constraint, a ratio of variance components, which results in precisely one optimal non-integer solution. For the two-facet example, rounding-off the optimal non-integer solution yields a solution equal to one of the optimal solutions, solution $(\hat{n}_1, \hat{n}_2)$ = (8,3), obtained by our method. Depending on the rounding-off, however, the optimal solutions of the two methods can differ as was demonstrated for the two-facet example in the introduction with $\rho^2 = .79$.

It has been shown that various constraints can be employed with the method proposed here. In general, the employment of equality constraints is discouraged because they will often lead to non-integer and/or non-optimal solutions. Woodward and Joe's method is an example where an equality constraint is used to specify an acceptable generalizability coefficient. With our method, employing in the two-facet example equality constraint $n_1 = 4n_2$ as an economic constraint and specifying $g \geq .80$ will result in solution $(\hat{n}_1, \hat{n}_2)$ = (12,3) with $L$ = 36 observations. However, $g \geq .80$ is also satisfied by solution $(\hat{n}_1, \hat{n}_2)$ = (8,3) with $L$ = 24 observations. The latter solution is of course to be preferred because it needs four fewer conditions for Facet 1. Employing more than one equality constraint will often even result in no integer solution at all, like when employing the two equality constraints $n_1 = 4n_2$ and $g = .80$ in the two-facet example. With respect to the number of constraints employed, it should be clear that adding constraints will reduce the set of feasible solutions.

The versatility of our method makes it extremely useful for investigators planning generalizability studies and practitioners involved with making decisions about the composition of measurement instruments. A computer program for the designs discussed here as well as other designs has been developed. Using as input $G$-study estimates

of variance components obtainable from existing computer programs for generalizability studies, for example, GENOVA (Crick & Brennan, 1982), computation time is no more than a few seconds.

The foregoing presentation has emphasized the practical aspects of the method proposed here. The method does however have important theoretical aspects as well. It can easily be seen that (2), the Spearman-Brown Prophecy formula from classical test theory, can be stated as a non-integer optimization problem with $L = n_1$ and $\rho^2(\hat{n}_1) = g$. By employing a threshold constraint and an integer constraint, this formula can be stated as an integer optimization problem. The search-tree of this problem consists of a root-node with one branch going down to one node at Level 1, which is associated with the only feasible value that minimizes $L$. The method proposed in this article generalizes the Spearman-Brown formula for measurement instruments with one facet to measurement instruments with more than one facet. As generalizability theory is the theoretical framework for these instruments, this method is a theoretical contribution to this framework.

Thus far the number of applications of integer optimization techniques to solve problems in psychometrics has been limited. Theunissen (1985) was the first to show how integer optimization techniques could be used to solve problems in latent trait theory. For generalizability theory these techniques also appear ultimately suited to handle a broad range of practical problems. Decision studies no longer have to be conducted through successive approximations (Cardinet & Tourneur, 1985). As a result of optimization techniques the decision-study phase of generalizability theory will no longer be characterized by trial and error but will be a phase characterized by rational decision making. More fruitful applications of these techniques are to be expected in the near future (e.g. van der Linden & Boekkooi-Timminga, 1989).

## References


Brennan, R.L. (1983). *Elements of generalizability theory.* Iowa City: ACT.

Cardinet, J. and Allal, L. (1983). Estimation of generalizability parameters. In: Leslie J. Fyans (Ed.), *Generalizability theory: Inferences and practical applications* (pp. 17-48). San Francisco: Jossey-Bass.

Cardinet, J. and Tourneur, Y. (1985). *Assurer la mesure* [Adequate Measurement]. Bern: Peter Lang.

Crick, J.E. and Brennan, R.L. (1982). *GENOVA: A generalized* analysis of variance system. (FORTRAN IV computer program and manual). IOWA City: ACT.

Cronbach, L.J., Gleser, G.C., Nanda, H. and Rajaratnam, N. (1972). *The dependability of behavioral measurements.* New York: Wiley.

Papadimitriou, Ch.H. and Steiglitz, K. (1982). *Combinatorial optimization: Algorithms and complexity.* Englewood Cliffs: Prentice-Hall.

Salkin, H.M. (1975). *Integer programming.* Reading: Addison-Wesley.

Theunissen, T.J.J.M. (1985). Binary programming and test design. *Psychometrika, 50,* 411-420.

van der Linden, W.J. and Boekkooi-Timminga, E. (1988). A maximin model for test design with practical constraints. *Psychometrika,* 54, 237-247.

Woodward, J.A. and Joe, G.W. (1973). Maximizing the coefficient of generalizability. *Psychometrika,* 38, 173-181.

# Chapter 3

## Maximizing the Coefficient of Generalizability Under the Constraint of Limited Resources*

* The material in this chapter has been published in Sanders, P.F., Theunissen, T.J.J.M., and Baas, S.M. (1991). Maximizing the coefficient of generalizability under the constraint of limited resources. *Psychometrika, 56,* 87-96.

## 3.1 Introduction

In generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) a distinction is made between a generalizability study and a decision study. In a generalizability study, accurate estimates of variance components are obtained that can be used in decision studies. Variance components from a generalizability study can, for example, be used to gauge the change in generalizability coefficients when altering the number of conditions for the facets in a decision study using the expression given below in (1). Explicitly, the generalizability coefficient, $\rho^2$, for the two-facet random-model crossed design is expressed as

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2} + \frac{\sigma_{res}^2}{n_1 n_2}}, \qquad (1)$$

where $\sigma_p^2$ is the variance component for persons, $\sigma_{p1}^2$ is the variance component for the person by Facet 1 interaction, $\sigma_{p2}^2$ is the variance component for the person by Facet 2 interaction, $\sigma_{res}^2$ is the variance component for the person x Facet 1 x Facet 2 interaction plus error, and $n_1$ and $n_2$ are the number of conditions of Facet 1 and Facet 2. The three interaction variance components constitute the error variance in the two-facet random-model crossed design. The product of the number of conditions of Facet 1 and Facet 2, $n_1 n_2$, forms the number of observations per subject or other object of measurement. From (1), an investigator or test constructor can not only infer that increasing the number of conditions of the two facets will increase the generalizability coefficient, but also that altering the number of conditions of a facet with a large error variance component will have a different impact on the generalizability coefficient than altering the number of conditions of a facet with a small error variance component.

To improve the quality and economy of decision studies, Woodward and Joe (1973) developed two procedures. One can be used to solve the problem of determining the minimum number of observations for a specified generalizability coefficient. Recently, a more versatile extension of their procedure for this problem was presented by Sanders, Theunissen, and Baas (1989). In response to discussions by Cronbach et al. (1972, pp. 173, 182), Woodward and Joe (1973) also developed a procedure for the problem of maximizing the generalizability coefficient subject to the constraint that the total number of observations is fixed. Because this constraint implies that the investigator is free to vary the number of conditions of the facets and resources required per condition of each facet do not differ, their procedure does not correspond to the application of decision studies in practice. Investigators do not conduct decision studies under the constraint of a fixed total number of observations but under other, different constraints, the most important being an upper limit on resources - including monetary expenditures, time, and expertise. Therefore, resources constitute a factor that has to be considered when decisions about the composition of measurement instruments are made. However, despite its relevance, procedures that could help investigators make these decisions are lacking.

The problem of maximizing the coefficient of generalizability under the constraint of an upper limit on resources is addressed in this paper. The problem is first formulated as an optimization problem. A two-step procedure for solving the optimization problem for the two-facet random-model crossed design is presented, and an example used to illustrate the procedure.

### 3.2 Optimization Problem

Maximizing the generalizability coefficient of the two-facet random-model crossed design is equivalent to minimizing the error variance. Denoting $\sigma^2_{p1}$, $\sigma^2_{p2}$ and $\sigma^2_{res}$ by $V_1$, $V_2$, and $V_{12}$ respectively, the optimization problem is formulated as

$$\text{minimize} \quad \frac{V_1}{n_1} + \frac{V_2}{n_2} + \frac{V_{12}}{n_1 n_2}. \tag{2}$$

The minimization statement in (2) refers to the value of the objective-function that results when different numbers of conditions, $n_1$ and $n_2$, are used for Facet 1 and 2.

A complete description of our optimization problem includes two other constraints. First, the constraint specifying the resources available for the decision study

$$c_1 n_1 + c_2 n_2 + c_{12} n_1 n_2 \leq r. \tag{3}$$

In (3), the resources required by the conditions of Facet 1 are specified by the term $c_1 n_1$, with $c_1$ being the 'cost' of one condition of Facet 1. The resources required by the conditions of Facet 2 are specified by the term $c_2 n_2$, with $c_2$ being the cost of one condition of Facet 2. The number of observations per subject or other object of measurement equals $n_1 n_2$. Thus, denoting the cost of one observation for the sample of subjects to be tested by $c_{12}$, the resources necessary for the total number of observations are specified by the term $c_{12} n_1 n_2$. The right-hand term of (3) specifies an upper limit on the resources available for the decision study. Here, it is assumed that the cost per condition does not vary within the same facet but can vary across different facets, and that the cost of conditions for all the facets are expressed in the same units as $r$. Second, because feasible values for $n_1$ and $n_2$ have to be integer and each facet in a

two-facet design has to have at least one condition, a lower bound integer constraint has to be included as well

$n_1$ and $n_2$ integer $\geq 1$. (4)

In addition to (3) and (4), other constraints can be employed, as discussed later.

### 3.3 Solution for the Two-Facet Design Problem

Optimal integer solutions for $n_1$ and $n_2$ of the optimization problem defined by (2), (3), and (4) are obtained in two steps. In the first, solutions are derived for a continuous relaxation of constraint (4), that is, $n_1, n_2 \geq 0$. In the second step, these optimal continuous solutions are used as the bounds in a branch-and-bound procedure (see Papadimitriou & Steiglitz, 1982, p. 443) to obtain the optimal integer solutions.

#### 3.3.1 Continuous Solution

By defining $\frac{1}{n_1} = x_1 + \frac{c_1}{r}$, $\frac{1}{n_2} = x_2 + \frac{c_2}{r}$, $w_{12} = v_{12}$, $w_1 = v_1 + \frac{v_{12}c_2}{r}$, and $w_2 = v_2 + \frac{v_{12}c_1}{r}$, a transformed version of the problem can be formulated as:

$$\text{minimize} \quad w_1x_1 + w_2x_2 + w_{12}x_1x_2 + \frac{v_1c_1}{r} + \frac{v_2c_2}{r} + \frac{v_{12}c_1c_2}{r^2} \tag{5}$$

$$\text{subject to} \quad x_1x_2 \geq \frac{c_{12}}{r} + \frac{c_1c_2}{r^2}, \tag{6}$$

$$-\frac{c_1}{r} \leq x_1 \leq 1 - \frac{c_1}{r}, \text{ and} \tag{7}$$

$$-\frac{c_2}{r} \leq x_2 \leq 1 - \frac{c_2}{r}.$$

However, $x_1 = \frac{1}{n_1} - \frac{c_1}{r} < 0$ or $x_2 = \frac{1}{n_2} - \frac{c_2}{r} < 0$ would imply $n_1 c_1 > r$ or $n_2 c_2 > r$, which is inconsistent with the resources constraint (3). Hence (7) is modified to $0 \le x_1 \le 1 - \frac{c_1}{r}$, and $0 \le x_2 \le 1 - \frac{c_2}{r}$, thus leading to bounds on $n_1$ of $(1, \frac{r}{c_1})$ and on $n_2$ of $(1, \frac{r}{c_2})$.

The formulation of the problem is further simplified by defining the right-hand terms of (6) and (7) as

$$b = \frac{c_{12}}{r} + \frac{c_1 c_2}{r^2}, \; u_1 = 1 - \frac{c_1}{r}, \text{ and } u_2 = 1 - \frac{c_2}{r},$$

so that the continuous optimization problem can be expressed as

$$\text{minimize} \quad w_1 x_1 + w_2 x_2 + w_{12} x_1 x_2 + \text{constant} \tag{8}$$

$$\text{subject to} \quad x_1 x_2 \ge b, \tag{9}$$

$$0 \le x_1 \le u_1, \text{ and } 0 \le x_2 \le u_2. \tag{10}$$

Note that the last three terms of (5) are denoted as a constant in (8). The above optimization problem can be solved by standard methods from non-linear optimization using Kuhn-Tucker necessary conditions. The Lagrange function is given by

$$L(X, \lambda) = w_1 x_1 + w_2 x_2 + w_{12} x_1 x_2 - \lambda_1 (x_1 x_2 - b) + \lambda_2 (x_1 - u_1)$$
$$- \lambda_3 x_1 + \lambda_4 (x_2 - u_2) - \lambda_5 x_2 .$$

To guarantee a nonempty solution set (see discussion later on), the assumption $u_1 u_2 \ge b$ or its equivalent $c_1 + c_2 + c_{12} \le r$ has to be made. Since a so-called constraint qualification holds with this assumption, the Lagrange function results in the following system of Kuhn-Tucker necessary conditions:

$$\frac{\partial L}{\partial x_1} = w_1 + w_{12} x_2 - \lambda_1 x_2 + \lambda_2 - \lambda_3 = 0; \tag{11}$$

$$\frac{\partial L}{\partial x_2} = w_2 + w_{12} x_1 - \lambda_1 x_1 + \lambda_4 - \lambda_5 = 0; \tag{12}$$

$$\lambda_1(x_1x_2 - b) = 0; \tag{13}$$

$$\lambda_2(x_1 - u_1) = 0; \tag{14}$$

$$\lambda_3 x_1 = 0; \tag{15}$$

$$\lambda_4(x_2 - u_2) = 0; \tag{16}$$

$$\lambda_5 x_2 = 0; \tag{17}$$

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0; \tag{18}$$

$$x_1x_2 \geq b; \tag{19}$$

$$0 \leq x_1 \leq u_1, \quad 0 \leq x_2 \leq u_2. \tag{20}$$

These conditions are necessary but not sufficient to produce a solution for the continuous problem being considered here. However, since the feasible region defined by (9) and (10) is closed and bounded in $R^2$, and the objective-function is continuous there, the existence of a solution is guaranteed. Moreover, any solution to the constrained minimization problem is included in the solution set for the Kuhn-Tucker conditions. This solution set will be derived next, with a candidate solution denoted by $x_1^*$ and $x_2^*$.

If either $x_1 = 0$ or $x_2 = 0$, (19) cannot be fulfilled since $b > 0$. Hence, in a feasible solution both $x_1 > 0$ and $x_2 > 0$, must hold, from which, because of (15) and (17), it follows that $\lambda_3 = \lambda_5 = 0$. Since $\lambda_2 \geq 0$ and $\lambda_4 \geq 0$, it is concluded that $\lambda_1 > 0$, otherwise (11) and (12) cannot have a solution. With $\lambda_1 > 0$, (13) implies that $x_1x_2 = b$ must hold.

A solution with $x_1^* = u_1$ and $x_2^* < u_2$, given $x_1^*x_2^* = b$, can only satisfy the Kuhn-Tucker conditions if $\frac{b}{u_1} < u_2$. Substituting these values in (11) and (12) with $\lambda_3 = \lambda_5 = 0$, as established above, and $\lambda_4 = 0$, because of (16), results in the following system of two

equations in the two variables $\lambda_1$ and $\lambda_2$ : $w_1 + w_{12}\frac{b}{u_1} - \lambda_1\frac{b}{u_1} + \lambda_2 = 0$, and $w_2 + w_{12}u_1 - \lambda_1 u_1 = 0$. Solving this system gives $\lambda_1 = \frac{w_2 + w_{12}u_1}{u_1}$, and $\lambda_2 = \frac{bw_2}{u_1^2} - w_1$. Since, because of (18), $\lambda_2 \geq 0$, a solution $x_1^* = u_1, x_2^* = \frac{b}{u_1} < u_2$ exists, if and only if $w_1 \leq \frac{bw_2}{u_1^2}$. Similarly, if $w_2 \leq \frac{bw_1}{u_2^2}$, solution $x_1^* = \frac{b}{u_2} < u_1, x_2^* = u_2$ results. A solution with $x_1^* = u_1, x_2^* = u_2$ ($\equiv n_1 = n_2 = 1$) can only exist if $u_1 u_2 = b$ ($\equiv c_1 + c_2 + c_{12} = r$), which has already been excluded from consideration.

What remains to be derived is the solution for the general case with $x_1^* < u_1$ and $x_2^* < u_2$, for which $\lambda_2 = \lambda_4 = 0$ according to (14) and (16). From (11), with $\lambda_2 = 0$, $x_2^* = \frac{w_1}{\lambda_1 - w_{12}}$; and from (12), with $\lambda_4 = 0$, $x_1^* = \frac{w_2}{\lambda_1 - w_{12}}$. These two equations give $x_2^* = \frac{w_1}{w_2} x_1^*$. Substitution in (9) produces $(x_1^{*2})^2 = b\frac{w_1}{w_2}$, which in turn yields $x_1^* = \left(\frac{bw_2}{w_1}\right)^{1/4}$, and $x_2^* = \left(\frac{bw_1}{w_2}\right)^{1/4}$.

It is easily verified that the four possible solutions exclude each other. For example, for the solution of the general case, solution $x_1^* = \left(\frac{bw_2}{w_1}\right)^{1/4} < u_1$ and $x_2^* = \left(\frac{bw_1}{w_2}\right)^{1/4} < u_2$ implies that the inequalities $w_1 > \frac{bw_2^2}{u_1^2}$ and $w_2 > \frac{bw_1}{u_2^2}$ must hold, which excludes as possible solutions $x_1^* = u_1$ and $x_2^* = u_2$. The four possible solutions are listed in Table 1. The last column of this table contains the solutions after backtransformation. Because solutions 2 and 3 with $n_1^*$ or $n_2^*$ equal to 1 are not likely to occur, the solution for the general case, solution 4, can be regarded as the appropriate optimal continuous solution for the two-facet optimization problem.

**TABLE 1**

**Four Possible Solutions for the Continuous Optimization Problem**

| | | | |
|---|---|---|---|
| 1. $u_1u_2 = b$ | $x_1^* = u_1$ | $\rightarrow$ | $n_1^* = 1$ |
| | $x_2^* = u_2$ | $\rightarrow$ | $n_2^* = 1$ |
| 2. $u_1u_2 > b$ | $x_1^* = u_1$ | $\rightarrow$ | $n_2^* = 1$ |
| $w_1 \leq \frac{bw_2}{u_1^2}$ | $x_2^* = \frac{b}{u_1}$ | $\rightarrow$ | $n_2^* = \left[\frac{rc_{12} + c_1c_2}{r(c-c_1)} + \frac{c_2}{r}\right]^{-1}$ |
| 3. $u_1u_2 > b$ | $x_1^* = \frac{b}{u_2}$ | $\rightarrow$ | $n_1^* = \left[\frac{rc_{12} + c_1c_2}{r(r-c_2)} + \frac{c_1}{r}\right]^{-1}$ |
| $w_1 \leq \frac{bw_2}{u_2^2}$ | $x_2^* = u_2$ | $\rightarrow$ | $n_2^* = 1$ |
| 4. $u_1u_2 > b$ | $x_1^* = \left[\frac{bw_2}{w_1}\right]^{1/3}$ | $\rightarrow$ | $n_1^* = \left[\left(\frac{w_2(rc_{12} + c_1c_2)}{w_1r^2}\right)^{1/3} + \frac{c_1}{r}\right]^{-1}$ |
| $w_1 > \frac{bw_2}{u_1^2}$, $w_2 > \frac{bw_1}{u_2^2}$ | $x_2^* = \left[\frac{bw_1}{w_2}\right]^{1/3}$ | $\rightarrow$ | $n_2^* = \left[\left(\frac{w_1(rc_{12} + c_1c_2)}{w_2r^2}\right)^{1/3} + \frac{c_2}{r}\right]^{-1}$ |

### 3.3.2 Integer Solution

The optimal continuous solution for the general case, derived in the preceding section, will be used as a starting point for a branch-and-bound procedure to obtain the optimal integer solution for the problem. Assume that $n_1^*$ and $n_2^*$ are the optimal continuous solutions, and both values are not integer. If $\lfloor n_1^* \rfloor$ denotes $n_1^*$ rounded down, and $\lceil n_1^* \rceil$ denotes $n_1^*$ rounded up, two problems have to be solved excluding the value $n_1^*$, but maintaining feasibility for all possible integer values of $n_1$. The first is

$$\text{minimize} \quad \frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2} + \frac{\sigma_{res}^2}{n_1 n_2}$$

$$\text{subject to} \quad c_1 n_1 + c_2 n_2 + c_{12} n_1 n_2 \leq r,$$

$$n_1, n_2 \geq 1, \text{ and}$$

$$n_1 \leq \lfloor n_1^* \rfloor.$$

In a solution $(\hat{n}_1, \hat{n}_2)$ for this problem, $\hat{n}_1 = \lfloor n_1^* \rfloor$ must hold, because if $n_1^* \neq \lfloor \hat{n}_1 \rfloor$, the rounding down constraint would be irrelevant, resulting in the non-integer solution to the original problem. As before, the resources constraint has to be satisfied as well. Consequently, either the solution for the first problem is $\hat{n}_1 = \lfloor n_1^* \rfloor$, and from the resources constraint, $\hat{n}_2 = \frac{r - c_1 \hat{n}_1}{c_2 + c_{12} \hat{n}_1}$, or $\hat{n}_1 = 1$ and $\hat{n}_2 = \frac{r - c_1}{c_2 + c_{12}}$. The second problem differs from the first in that the rounding down constraint is replaced by the rounding up constraint $n_1 \geq \lceil n_1^* \rceil$. The solution for the second problem is $\hat{n}_1 = \lceil n_1^* \rceil$, and $\hat{n}_2 = \frac{r - c_1 \hat{n}_1}{c_2 + c_{12} \hat{n}_1}$.

Each of these two problems gives rise to two new problems, with additional constraints regarding $n_2$ whenever the value $\hat{n}_2$ is non-integer. If, for instance, $\hat{n}_2$ is non-integer in the first problem, the constraint sets for the following two problems are given by (1) $n_2 \geq 1$, $1 \leq n_1 \leq \lfloor \hat{n}_1 \rfloor$, and (2) $n_2 \geq 1$, $n_1 \geq \lceil \hat{n}_1 \rceil$. Branching continues until either an integer solution or an infeasible constraint set is found. Each time an integer solution is obtained during the branching process, its value is compared with that of the previous best solution and accepted as the new best solution or rejected as such. At the end of the process, the current best solution is the optimal integer solution.

### 3.4 Example

Our procedure will be illustrated for the two-facet random-model crossed design using the example from Woodward and Joe (1973, p. 179) with the following variance components: $\hat{\sigma}^2_p = 5.435$, $\hat{\sigma}^2_{p1} = 3.421$, $\hat{\sigma}^2_{p2} = 1.140$, and $\hat{\sigma}^2_{res} = 11.850$. The objective-function for this example is

$$\text{minimize} \quad \frac{3.421}{n_1} + \frac{1.140}{n_2} + \frac{11.850}{n_1 n_2}.$$

Assuming that a condition of Facet 1 (e.g., essay questions), costs 40 dollars, a condition of Facet 2 (e.g., raters), costs nothing and one observation for all the subjects in the sample (i.e., the answers of all examinees to one essay question rated by one rater), costs 80 dollars, and that the budget for the decision study is limited to 3000 dollars, the resources constraint for this example can be stated as

$$40n_1 + 80n_1n_2 \leq 3000.$$

The optimal integer solutions $\hat{n}_1$ and $\hat{n}_2$ for this optimization problem are derived in two steps. First, using solution 4 from Table 1, the optimal continuous solutions $n_1^* = 8.8$ and $n_2^* = 3.8$ are obtained. Second, because both solutions are non-integer, a branch-and-bound procedure is needed to find the optimal integer solutions. For the problem with additional constraint $n_1 \leq \lfloor n_1^* \rfloor = 8$, one finds a solution $\hat{n}_1 = 8$, $\hat{n}_2 = 4.2$ and for the problem with $n_1 \geq \lceil n_1^* \rceil = 9$, a solution $\hat{n}_1 = 9$, $\hat{n}_2 = 3.7$. Further branching gives rise to four problems with constraint sets including $n_1, n_2 \geq 1$ and: (1) $n_1 \leq 8$, $n_2 \leq 4$; (2) $n_1 \leq 8$, $n_2 \leq 5$; (3) $n_1 \leq 9$, $n_2 \leq 3$, and (4) $n_1 \leq 9$, $n_2 \leq 4$. The search-process for the example with resources limited to 3000 dollars is shown in Figure 1.

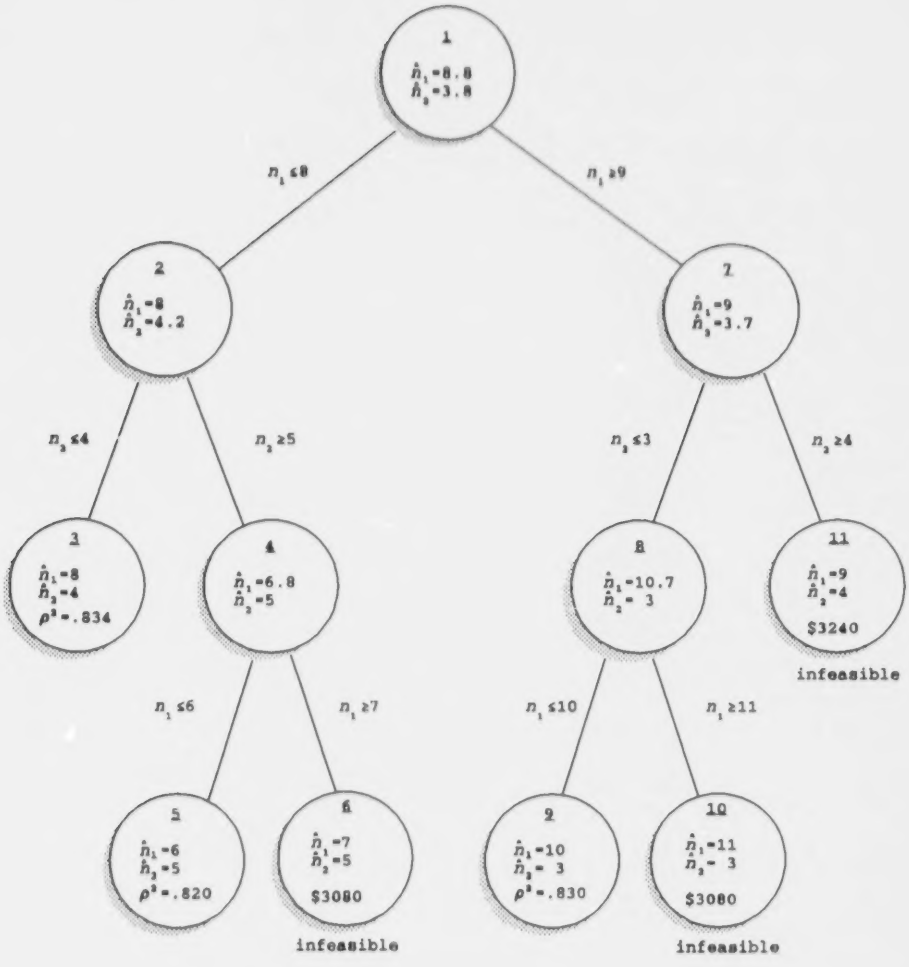


**FIGURE 1**

Search-tree for the two-facet example

Starting from the optimal continuous solution in Node 1, the strategy to traverse the search-tree is depth-first and from left to right, as indicated by the numbering of the nodes. Node 3 produces the first candidate solution, which is not improved by the solution found in Node 5. The solution in Node 6 does not satisfy the budget constraint of 3000 dollars and is therefore an infeasible solution. Other infeasible solutions are found in Node 10 and 11. The solution found in Node 9 is slightly worse than the solution in Node 3 and thus

rejected. From the search-process in Figure 1 as well other solutions presented in Table 2, it can be concluded that a more exhaustive search will result in either infeasible solutions or solutions that do not improve the solution in Node 3. Therefore, the solution in Node 3, $(\hat{n}_1, \hat{n}_2) = (8,4)$, that is, eight essay questions of which the answers of all examinees have to be rated by four raters, is the optimal integer solution for this problem.

**TABLE 2**

Values of $n_1$, $n_2$, Variance Components, $\rho^2$ when $c_1$ = 40 Dollars, $\rho^2$ = 40 Dollars, $c_2$ = 0 Dollars and $c_{12}$ = 80 Dollars

| $n_1$ | $c_1 n_1$ | $n_2$ | $n_1 n_2$ | $c_{12} n_1 n_2$ | $\hat{\sigma}^2_p$ | $\frac{\hat{\sigma}^2_{p1}}{n_1}$ | $\frac{\hat{\sigma}^2_{p2}}{n_2}$ | $\frac{\hat{\sigma}^2_{res}}{n_1 n_2}$ | $\rho^2$ | $r$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 240 | 4 | 24 | 1920 | 5.435 | .570 | .285 | .494 | .802 | 2160 |
| 6 | 240 | 5 | 30 | 2400 | 5.435 | .570 | .228 | .395 | .820 | 2640 |
| 7 | 280 | 3 | 21 | 1680 | 5.435 | .489 | .380 | .564 | .791 | 1960 |
| 7 | 280 | 4 | 28 | 2240 | 5.435 | .489 | .285 | .423 | .820 | 2520 |
| 7 | 280 | 5 | 35 | 2800 | 5.435 | .489 | .228 | .339 | .837 | 3080 |
| 8 | 320 | 3 | 24 | 1920 | 5.435 | .428 | .380 | .494 | .807 | 2240 |
| 8 | 320 | 4 | 32 | 2560 | 5.435 | .428 | .285 | .370 | .834 | 2880 |
| 9 | 360 | 3 | 27 | 2160 | 5.435 | .380 | .380 | .439 | .819 | 2520 |
| 9 | 360 | 4 | 36 | 2880 | 5.435 | .380 | .285 | .329 | .845 | 3240 |
| 10 | 400 | 3 | 30 | 2400 | 5.435 | .342 | .380 | .395 | .830 | 2800 |
| 10 | 400 | 4 | 40 | 3200 | 5.435 | .342 | .285 | .296 | .855 | 3600 |
| 11 | 440 | 3 | 33 | 2640 | 5.435 | .311 | .380 | .359 | .838 | 3080 |
| 12 | 480 | 2 | 24 | 1920 | 5.435 | .285 | .570 | .494 | .801 | 2400 |

Although the resources constraint employed here is not based on real data, the example is realistic since in most decision studies, differences between the number of observations will have more impact on the necessary resources than on the generalizability coefficient. A small difference between generalizability coefficients will therefore correspond with a large difference in resources (for example, compare solution $(\hat{n}_1, \hat{n}_2) = (7,3)$ with solution $(\hat{n}_1, \hat{n}_2) = (8,3)$). Note that while the number of subjects employed in a decision study does not affect the generalizability coefficient directly, it can affect the

coefficient through $c_{12}$. Depending on the value for $c_{12}$ the optimal values of $n_1, n_2$, and therefore $\rho^2$, can change considerably.

The present example is also used to demonstrate that our procedure is a generalization of that of Woodward and Joe. Their procedure is a simplification of the resources constraint (9) employed by our procedure, obtained by substituting the left-hand term of (9) with the number of observations, the 'greater than or equal to' sign with the 'equal' sign and the right-hand term with a specific number of observations. Then, using solution 4 from Table 1, the optimal continuous number of conditions of Facet 1 and Facet 2 under the constraint of a fixed number of observations is given by

$$n_1^* = \left[\frac{b^{-1}w_1}{w_2}\right]^{1/2}, \quad \text{and} \quad n_2^* = \left[\frac{b^{-1}w_2}{w_1}\right]^{1/2}.$$

Substituting $b^{-1}$, the number of observations specified by the investigator, with the term L, leads to the Woodward and Joe equations (1973, p. 176). According to the preceding equations, the optimal continuous number of conditions of Facet 1 and Facet 2, when L is equal to 400, are $n_1^* = 34.65$ and $n_2^* = 11.55$. Rounding off these non-integer values, their solution ($\hat{n}_1 = 40$, $\hat{n}_2 = 10$) with a generalizability coefficient equal to .9595, is the optimal integer solution under the constraint of 400 observations. However, a generalizability coefficient of .9596 can be obtained with a smaller number of observations, explicitly, $\hat{n}_1 = 36$ and $\hat{n}_2 = 11$. This solution would have been obtained with our two-step procedure, and it can be proven that this solution is indeed the optimal integer solution (Sanders et al., 1989). Although the differences in generalizability coefficients are trivial, this solution is to be preferred over solutions ($\hat{n}_1 = 33$, $\hat{n}_2 = 12$) and

($\hat{n}_1 = 35$, $\hat{n}_2 = 11$) with respectively 396 and 385 observations. The latter procedure also shows the insensitivity of higher values of generalizability coefficients to even major changes in the design. Woodward and Joe's solution is less versatile than ours because their procedure does not allow for the number of observations to vary. With L set to 396, their solution would have been the same as ours. Besides the efficiency resulting from defining the objective-function as an inequality constraint, our procedure also allows for handling additional constraints much more easily.

As was shown by the example above, it is important to note that when the resources constraint is replaced by a constraint specifying the number of observations, then $w_1 = v_1$ and $w_2 = v_2$ in the above expressions. For resources constraints comparable to the budget constraint employed in our example, the ratio $w_1/w_2$ will roughly approximate the ratio $v_1/v_2$.

### 3.5 Conclusions and Discussion

The procedure proposed in this article enables an investigator or test constructor to conduct a decision study under the constraint of limited resources. Although in practice, resources will be the most important constraint, the procedure admits the possibility of others as well. For example, an investigator who wishes a measurement instrument composed of at least six times as many conditions of Facet 1 (e.g., essay questions) than of Facet 2 (e.g., raters) can achieve this by employing the constraint $n_1 \geq 6n_2$. Adding this constraint to the present example would result in the optimal solution ($\hat{n}_1 = 12$, $\hat{n}_2 = 2$) = 2400 dollars, as can be seen in Table 2. The possibility of including

additional constraints makes the procedure particularly useful for solving many practical measurement construction problems.

Because of the consequences for the measurement instruments attached to the choice of constraints, they should be the result of a deliberate choice and carefully examined. Following Sanders et al. (1989), a distinction can be made between psychometric constraints and other, economic or practical, constraints. A psychometric constraint is one in accord with the psychometric structure of the problem - the magnitude of the error variance components. The inequality $n_1 \geq n_2$, for instance, would be a psychometric constraint for the present example, because a higher generalizability coefficient will be obtained by increasing the number of conditions of Facet 1 than by increasing the number of conditions of Facet 2. On the other hand, the constraint $n_1 \leq n_2$ would not be considered as such. Another useful distinction to be made is between equality and inequality constraints. Inequality constraints are to be preferred over equality constraints, because equality constraints often lead to less efficient solutions. The latter was illustrated by the Woodward and Joe equality constraint of a fixed number of observations. It should be emphasized that dependent on the employment of psychometric and/or other constraints, quite different optimal solutions can be obtained. Obviously, employing only psychometric constraints will lead to higher generalizability coefficients than employing other constraints. An investigator who considers the generalizability coefficient obtained inadequate, can take two decisions. The first is to change the resources constraint (e.g., increasing the budget). The second is to change the design of the decision study (e.g., nesting or deleting a facet).

An optimization model for classical test theory has recently been presented by van der Linden and Adema (1988). Assuming an item bank with estimates of item parameters based on classical test theory (i.e., item difficulty and item test correlation), their procedure maximizes coefficient alpha permitting as many practical constraints as test constructors can think of. Their procedure can be viewed as an extension of our procedure for one-facet designs. Instead of a (stratified) random selection of items from the item bank, their procedure makes a selection of specific items based on the psychometric properties of these items. For tests of usual length, however, the differences between reliability coefficients constructed by the two procedures can be expected to be small and of little practical relevance. For one-facet designs, the aspect of diminishing returns should also be taken into account, because higher values of reliability coefficients are hardly influenced by changes in the number of items.

Two extensions of the procedure presented here are obvious. First, the assumption of constant cost per condition within the same facet can be relaxed to allow for varying costs for conditions within the same facet. Second, the procedure can be generalized to multifacet and nested designs, and designs for mixed models. Although the mathematics of these generalizations is rather tedious, they are quite straightforward as demonstrated for a similar procedure by Sanders et al. (1989).

Applications of integer optimization techniques for test design within the framework of latent trait theory have been presented by Theunissen (1985), and van der Linden and Boekkooi-Timminga (1989). To make these techniques accessible to the practitioner, a computer program has been developed that can handle a variety of test designs

and constraints. Information about this program, called 'optimal test design', can be obtained from the second author of the current paper. Several computer programs for performing generalizability and decision studies (see Brennan, 1983; Cardinet & Tourneur, 1985; Crick & Brennan, 1982) have been developed. Unfortunately, these programs do not use optimization techniques. What is needed is a computer program that can handle various kinds of objective-functions, designs and complex constraints.

## References


Brennan, R.L. (1983). *Elements of generalizability theory*. Iowa City: ACT.

Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure [Adequate Measurement]*. Bern: Peter Lang.

Crick, J.E., & Brennan, R.L. (1982). *GENOVA: A generalized analysis of variance system*. (FORTRAN IV computer program and manual). Iowa City: ACT.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Papadimitriou, Ch.H., & Steiglitz, K. (1982). *Combinatorial optimization: Algorithms and complexity*. Englewood Cliffs: Prentice-Hall.

Sanders, P.F., Theunissen, T.J.J.M., & Baas, S.M. (1989). Minimizing the number of observations: A generalization of the Spearman-Brown formula. *Psychometrika, 54*, 587-598.

Theunissen, T.J.J.M. (1985). Binary programming and test design. *Psychometrika, 50*, 411-420.

van der Linden, W.J., & Boekkooi-Timminga, E. (1989). A maximin model for test design with practical constraints. *Psychometrika, 54*, 237-247.

van der Linden, W.J., & Adema, J.J. (1988). *Algorithmic test design using classical item parameters*. (Research Report 88-2). Enschede, The Netherlands: University of Twente, Department of Education.

Woodward, J.A., & Joe, G.W. (1973). Maximizing the coefficient of generalizability in multi-facet decision studies. *Psychometrika, 38*, 173-181.





BEST COPY AVAILABLE

# Chapter 4

# The Optimum Allocation of Measurements in Balanced Random Effects Models*

---

* The material in this chapter has been published in Sanders, P.F., and Eggen, T.J.H.M. (1991). *The optimum allocation of measurements in balanced random effects models*. Measurement and Research Department Reports (91-8). Arnhem: Cito.
This chapter has been submitted for publication as an article.

## 4.1 Introduction

Consider the balanced three-way random effects model with one observation per cell

$$Y_{ijk} = \mu + A_i + B_j + C_k + AB_{ij} + AC_{ik} + BC_{jk} + E_{ijk},$$
$$i = 1, \ldots, n_i,\ j = 1, \ldots, n_j,\ k = 1, \ldots, n_k,$$

where $\mu$ is a constant and the $\{A_i\}$, $\{B_j\}$, $\{C_k\}$, $\{AB_{ij}\}$, $\{AC_{ik}\}$, $\{BC_{jk}\}$, and $\{E_{ijk}\}$ are independent normals with zero means and variances $\sigma_A^2$, $\sigma_B^2$, $\sigma_C^2$, $\sigma_{AB}^2$, $\sigma_{AC}^2$, $\sigma_{BC}^2$, $\sigma_e^2$, respectively.

Parameters expressing the proportion of total variability are used in all fields of applications in which the variance component model is employed. In genetics, ratios of variance components are used as estimates of genetic heritability (e.g., Graybill & Wang 1979). In psychometrics, the variance components ratio

$$\rho_A = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_{AB}^2/n_j + \sigma_{AC}^2/n_k + \sigma_e^2/n_j n_k} \qquad (1)$$

is an important parameter (e.g., Cronbach, Gleser, Nanda & Rajaratnam 1972). This parameter is an indicator of the degree of precision provided by a measurement instrument in which factor $A$ constitutes the objects of measurement, that is, examinees, whereas factor $B$ and $C$ each constitute sets of conditions of measurement, for example, $n_j$ questions and $n_k$ raters. The number of measurements per examinee equals $n_j n_k$.

In (1), $\sigma_{AB}^2$, $\sigma_{AC}^2$ and $\sigma_e^2$ constitute the measurement error. From (1), it can be inferred that measurement precision can be improved by increasing $n_j$ and $n_k$. However, because in practice the administration of measurements has to be conducted under the constraint of limited resources, the optimal allocation of $n_j$ and $n_k$ is of critical importance. A procedure for maximizing $\rho_A$ under the constraint of a fixed number of measurements was first proposed by Woodward and Joe (1973). With this procedure, however, non-optimal solutions are likely

to occur because of the equality constraint on $n_j n_k$. Moreover, the generalization of Woodward and Joe's procedure as well as the procedure developed by Sanders, Theunissen, and Baas (1991), which does provide optimal solutions under a budget constraint, to four-way and other designs becomes very complex.

Addressing the same problem of maximizing $\rho_A$ under a budget constraint, an alternative procedure is presented in this paper. In section 4.2, the maximization problem of the three-way random effects model is formulated as an optimization problem. In section 4.3, a two-step procedure for solving this problem is proposed and illustrated. The solution for the maximization problem of the four-way random effects model is presented in section 4.4.

## 4.2 Optimization Problem

Maximizing (1) is equivalent to minimizing the objective-function

$$f(n_j, n_k) = \frac{\sigma^2_{AB}}{n_j} + \frac{\sigma^2_{AC}}{n_k} + \frac{\sigma^2_e}{n_j n_k} . \qquad (2)$$

Minimization statement (2) refers to the value of the objective-function with different values for $n_j$ and $n_k$.

A complete problem specification includes two other constraints. First, the constraint that specifies the available resources, that is, the budget. As costs in general vary in proportion to $n_j n_k$, the cost function is defined as

$$c_{jk}\, n_j\, n_k \leq C,$$

where $c_{jk}$ denotes the cost of one measurement, for example, the rating by one rater of the answers of all examinees in the sample to one question, and $C$ the budget. Assuming that the cost of one measurement is the same for each examinee, the cost function can be written as

$$n_j\, n_k \leq \frac{C}{C_{jk}} = M, \tag{3}$$

where the number of measurements is denoted by $M$. Second, because the nature of the problems under consideration demands that feasible values for $n_j$ and $n_k$ be integer values and that each factor has at least one condition of measurement, a lower bound integer constraint is specified

$$n_j \text{ and } n_k \text{ integer} \geq 1. \tag{4}$$

### 4.3 Proposed Solution Three-Way Design

Optimal integer solutions for $n_j$ and $n_k$ of the optimization problem defined by (2), (3) and (4) are obtained in two steps. In the first, solutions are derived for a continuous relaxation of constraint (4), that is, $n_j,\ n_k > 0$. These optimal continuous solutions are used in the second step as the bounds in a branch-and-bound procedure (see Papadimitriou & Steiglitz 1982, p. 443).

#### 4.3.1 Continuous Solution

The continuous problem can be solved by using Lagrange multipliers method. The Lagrange function is given by

$$L\ (n_j, n_k, \lambda) = \frac{\sigma^2_{AB}}{n_j} + \frac{\sigma^2_{AC}}{n_k} + \frac{\sigma^2_e}{n_j n_k} + \lambda\,(n_j n_k - M)\,. \tag{5}$$

The necessary conditions for the minimum of $f(n_j, n_k)$ are

$$\frac{\delta L}{\delta n_j} = -\frac{\sigma^2_{AB}}{n_j^2} - \frac{\sigma^2_e}{n_j^2 n_k} + \lambda n_k = 0, \tag{6}$$

$$\frac{\delta L}{\delta n_k} = -\frac{\sigma^2_{AC}}{n_k^2} - \frac{\sigma^2_e}{n_j n_k^2} + \lambda n_j = 0, \tag{7}$$

$$\frac{\delta L}{\delta \lambda} = n_j n_k - M = 0. \qquad (8)$$

From (6), (7) and (8) the optimal continuous solutions are derived as

$$n_j^* = \left( M \frac{\sigma_{AB}^2}{\sigma_{AC}^2} \right)^{1/2}, \text{ and} \qquad (9)$$

$$n_k^* = \left( M \frac{\sigma_{AC}^2}{\sigma_{AB}^2} \right)^{1/2}. \qquad (10)$$

Note that because of constraint (3), the minimization of (2) without the error term will give the same solutions.

#### 4.3.2 Integer Solution

The optimal continuous solutions, derived in the preceding section, are used as a starting point for a branch-and-bound algorithm to obtain the optimal integer solutions. Assume that $n_j^*$ and $n_k^*$ are the optimal continuous solutions, and both values are not integer. If $\lfloor n_j^* \rfloor$ denotes $n_j^*$ rounded down and $\lceil n_j^* \rceil$ denotes $n_j^*$ rounded up, two problems have to be solved. The first problem is

$$\text{minimize } f(n_j, n_k) = \frac{\sigma_{AB}^2}{n_j} + \frac{\sigma_{AC}^2}{n_k} + \frac{\sigma_e^2}{n_j n_k},$$

subject to $n_j\, n_k \le M$,

$n_j,\ n_k \ge 1$, and

$n_j \le \lfloor n_j^* \rfloor$.

In a solution $(\hat{n}_j,\ \hat{n}_k)$ for this problem, $\hat{n}_j = \lfloor n_j^* \rfloor$ must hold, because if $\hat{n}_j \neq \lfloor n_j^* \rfloor$, the rounding down constraint would be irrelevant

and a non-integer solution would result. Hence, the solution for the first problem is $\hat{n}_j = \lfloor n_j^* \rfloor$ and $\hat{n}_k = M/\hat{n}_j$, or $\hat{n}_j = 1$ and $\hat{n}_k = M$. The second problem differs from the first in that the rounding down constraint is replaced by the rounding up constraint $n_j \geq \lceil n_j^* \rceil$. The solution for the second problem is $\hat{n}_j = \lceil n_j^* \rceil$ and $\hat{n}_k = M/\hat{n}_j$. Each of these two problems yield two new problems with additional constraints regarding $n_k$ whenever the value $\hat{n}_k$ is non-integer. If, for instance, $\hat{n}_k$ is non-integer in the first problem, the constraint sets for the following two problems are given by (1) $n_k \geq 1$, $n_j \geq \lceil \hat{n}_j \rceil$, and (2) $n_k \geq 1$, $1 \leq n_j \leq \lfloor \hat{n}_j \rfloor$. Branching continues until either an integer solution or an infeasible constraint set is found. Each time an integer solution is obtained during the branching process, its value is compared with that of the previous best solution and accepted as the new best solution or rejected as such. At the end of the process, the current best solution is the optimal integer solution.

#### 4.3.3 Example

For the balanced three-way random effects model crossed design the example from Sanders et al. (1991 ) with variance components $\hat{\sigma}_A^2 = 5.435$, $\hat{\sigma}_{AB}^2 = 3.421$, $\hat{\sigma}_{AC}^2 = 1.140$, and $\hat{\sigma}_e^2 = 11.850$ is used. The objective-function for this example is

$$\text{minimize} \quad \frac{3.421}{n_j} + \frac{1.140}{n_k} + \frac{11.850}{n_j n_k} .$$

Assuming that conditions of factor $B$, for example, questions, and factor $C$, for example, raters, cost nothing, but that the rating by one rater of the answers of all examinees in the sample to one question costs 80 dollars, and that the budget is limited to 3000 dollars, the budget constraint for this example is stated as

$$80 n_j n_k \leq 3000. \qquad (11)$$

The optimal integer solutions $\hat{n}_j$ and $\hat{n}_k$ for this optimization problem are derived in two steps. First, using equations (9) and (10), the optimal continuous solutions $n_j^*$ = 10.6 and $n_k^*$ = 3.5 are obtained. Second, because both solutions are non-integer, a branch-and-bound algorithm is employed to obtain the optimal integer solutions. For the problem with additional constraint $n_j \leq \lfloor n_j^* \rfloor = 10$, one finds a solution $\hat{n}_j$ = 10, $\hat{n}_k$ = 3.7 and for the problem with $n_j \geq \lceil n_j^* \rceil = 11$, a solution $\hat{n}_j$ = 11, $\hat{n}_k$ = 3.4. Further branching yields four problems with constraint sets, including $n_j$, $n_k \geq 1$, (1) $n_j \leq 10$, $n_k \leq 3$; (2) $n_j \leq 10$, $n_k \geq 4$; (3) $n_j \geq 11$, $n_k \leq 3$, and (4) $n_j \geq 11$, $n_k \geq 4$. The search process is shown in Figure 1.

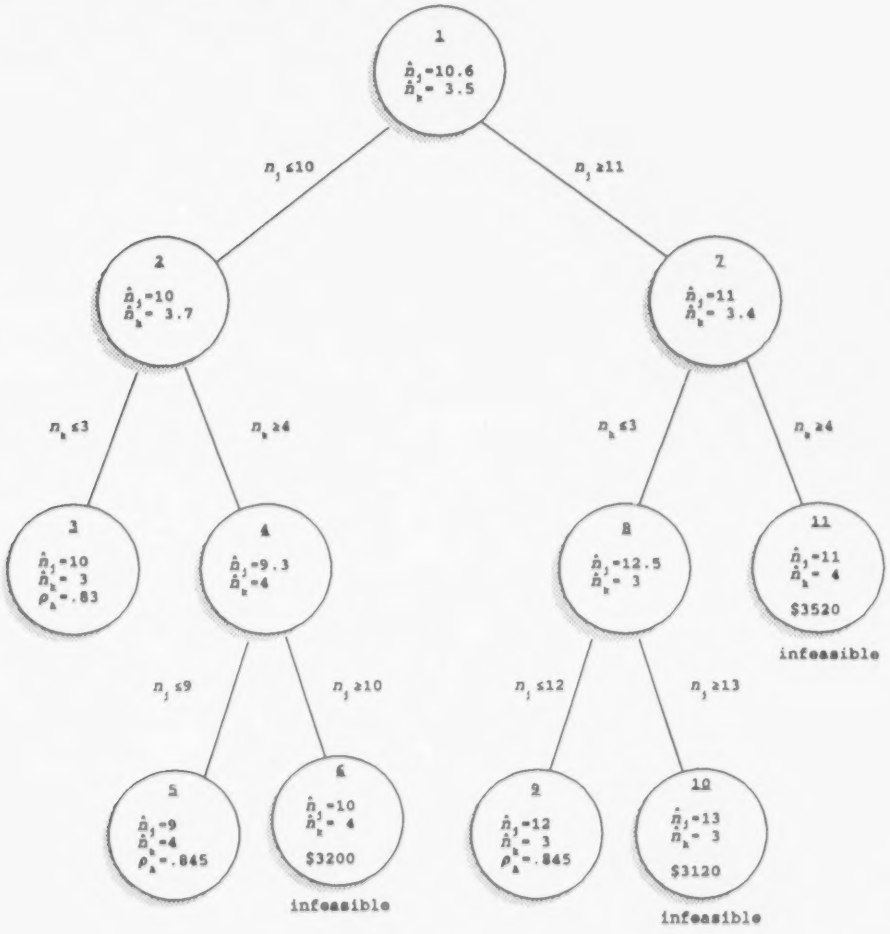


**FIGURE 1**

Search-Tree for the Three-Way Design Example

Starting from the optimal continuous solution in Node 1, the strategy to traverse the search-tree is depth-first and from left to right, as indicated by the numbering of the nodes. Node 3 produces the first candidate solution, which is improved by the solution found in Node 5. The solution in Node 6 does not satisfy the budget constraint of 3000 dollars and is therefore an infeasible solution. Other infeasible solutions are found in Node 10 and 11. The solution found in Node 9 is equal to the solution found in Node 5. From the search-process in Figure 1 and the solutions for this cost function presented in Table 1, it can be concluded that a more exhaustive search will result in either infeasible solutions or solutions that do not improve the solutions in Node 5 and 9. Therefore, the solution in Node 5, $(\hat{n}_j, \hat{n}_k) = (9,4)$, that is, nine questions for which the answers of all examinees have to be rated by four raters, or the solution in Node 9, $(\hat{n}_j, \hat{n}_k) = (12,3)$, that is twelve questions rated by three raters, are the optimal integer solutions for this problem.

TABLE 1

Values of $n_j$, $n_k$, Variance Components, $\rho_A$ and $C$ for Two Cost Functions

| $n_j$ | $c_j n_j$ | $n_k$ | $n_j n_k$ | $c_{jk} n_j n_k$ | $\hat{\sigma}^2_A$ | $\frac{\hat{\sigma}^2_{AB}}{n_j}$ | $\frac{\hat{\sigma}^2_{AC}}{n_k}$ | $\frac{\hat{\sigma}^2_e}{n_j n_k}$ | $\rho_A$ | $C$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. $c_{jk}$ = 80 Dollars, $C$ = 3000 Dollars | | | | | | | | | | |
| 10.6 | 0 | 3.5 | 37.5 | 3000 | 5.435 | .322 | .322 | .316 | .850 | 3000 |
| 9 | 0 | 4 | 36 | 2880 | 5.435 | .380 | .285 | .329 | .845 | 2880 |
| 11 | 0 | 4 | 44 | 3520 | 5.435 | .311 | .285 | .269 | .863 | 3520 |
| 10 | 0 | 3 | 30 | 2400 | 5.435 | .342 | .380 | .395 | .830 | 2400 |
| 11 | 0 | 3 | 33 | 2640 | 5.435 | .311 | .380 | .359 | .838 | 2640 |
| 10 | 0 | 4 | 40 | 3200 | 5.435 | .342 | .285 | .296 | .855 | 3200 |
| 12 | 0 | 3 | 36 | 2880 | 5.435 | .285 | .380 | .329 | .845 | 2880 |
| 13 | 0 | 3 | 39 | 3120 | 5.435 | .263 | .380 | .304 | .852 | 3120 |
| 2. $c_j$ = 40 Dollars, $c_{jk}$ = 80 Dollars, and $C$ = 3000 Dollars | | | | | | | | | | |
| 8 | 320 | 4 | 32 | 2560 | 5.435 | .428 | .285 | .370 | .834 | 2880 |
| 9 | 360 | 4 | 36 | 2880 | 5.435 | .380 | .285 | .329 | .845 | 3240 |
| 10 | 400 | 3 | 30 | 2400 | 5.435 | .342 | .380 | .395 | .830 | 2800 |
| 11 | 440 | 3 | 33 | 2640 | 5.435 | .311 | .380 | .359 | .838 | 3080 |

Although the budget constraint employed here was not taken from an actual measurement problem, the example is realistic since in most cases, differences between the number of measurements will have more impact on the budget than on $\rho_A$. This means that in many cases nearly the same measurement precision can be obtained for a considerably smaller budget. Table 1 also contains the solutions for a second cost function

$$40n_j + 80n_jn_k \leq 30000, \tag{12}$$

where $(c_j =)$ 40 dollars denotes the cost of one question. For all linear cost functions with the $n_jn_k$ measurements taking up the major part of the budget, the optimal continuous solution for a problem with cost function (11) can be used as a starting point of a branch-and-bound algorithm for a problem with cost function (12). The optimal continuous solution $\hat{n}_j$ for a problem with cost function (11) is the upper bound of $\hat{n}_j$ for a problem with cost function (12). It can be inferred from the results presented in Table 1, that a simple branch-and-bound algorithm will produce the optimal integer solution for problem (12), that is, eight questions rated by four raters.

### 4.4 Proposed Solution Four-Way Design

Consider the balanced four-way random effects model with one observation per cell

$$Y_{ijkm} = \mu + A_i + B_j + C_k + D_m + AB_{ij} + AC_{ik} + AD_{im} + BC_{jk} + BD_{jm} + CD_{km} + ABC_{ijk} + ABD_{ijm} + ACD_{ijm} + BCD_{jkm} + E_{ijkm},$$
$$i = 1,\ldots, n_i,\ j = 1,\ldots, n_j,\ k = 1,\ldots, n_k,\ m = 1,\ldots, n_m,$$

where $\mu$ is a constant and the $\{A_i\}$, $\{B_j\}$, $\{C_k\}$, $\{D_m\}$, $\{AB_{ij}\}$, $\{AC_{ik}\}$, $\{AD_{im}\}$, $\{BC_{jk}\}$, $\{BD_{jm}\}$, $\{CD_{km}\}$, $\{ABC_{ijk}\}$, $\{ABD_{ijm}\}$, $\{ACD_{ijm}\}$, $\{BCD_{jkm}\}$, and $\{E_{ijkm}\}$ are independent normals with zero means and variances

$\sigma^2_A$, $\sigma^2_B$, $\sigma^2_C$, $\sigma^2_D$, $\sigma^2_{AB}$, $\sigma^2_{AC}$, $\sigma^2_{AD}$, $\sigma^2_{BC}$, $\sigma^2_{BD}$, $\sigma^2_{CD}$, $\sigma^2_{ABC}$, $\sigma^2_{ABD}$, $\sigma^2_{ACD}$, $\sigma^2_{BCD}$, and $\sigma^2_e$, respectively. The maximization of the generalizability coefficient for the four-way design, amounts to minimizing

$$f(n_j n_k n_m) = \frac{\sigma^2_{AB}}{n_j} + \frac{\sigma^2_{AC}}{n_k} + \frac{\sigma^2_{AD}}{n_m} + \frac{\sigma^2_{ABC}}{n_j n_k} + \frac{\sigma^2_{ABD}}{n_j n_m} + \frac{\sigma^2_{ACD}}{n_k n_m} + \frac{\sigma^2_e}{n_j n_k n_m},$$

subject to $C_{jkm}\, n_j n_k n_m \leq C \rightarrow n_j n_k n_m \leq \frac{C}{C_{jkm}} = M.$

The application of the Lagrange multipliers method to this minimization problem results in the equations

$$n_m = n_k \left( \frac{\sigma^2_{AD} + \frac{\sigma^2_{ABD}}{n_j}}{\sigma^2_{AC} + \frac{\sigma^2_{ABC}}{n_j}} \right), \quad n_m = n_j \left( \frac{\sigma^2_{AD} + \frac{\sigma^2_{ACD}}{n_k}}{\sigma^2_{AB} + \frac{\sigma^2_{ABC}}{n_k}} \right), \text{ and}$$

$$n_k = n_j \left( \frac{\sigma^2_{AC} + \frac{\sigma^2_{ACD}}{n_m}}{\sigma^2_{AB} + \frac{\sigma^2_{ABD}}{n_m}} \right).$$

As second-order interaction variance components are generally very small, the above equations can be written as

$$\frac{n_m}{n_k} = \frac{\sigma^2_{AD}}{\sigma^2_{AC}}, \quad \frac{n_m}{n_j} = \frac{\sigma^2_{AD}}{\sigma^2_{AB}}, \text{ and } \quad \frac{n_k}{n_j} = \frac{\sigma^2_{AC}}{\sigma^2_{AB}}.$$

From these equations and the cost function, nearly optimal continuous solutions for $n_j$, $n_k$ and $n_m$ can be derived. These solutions are used as the starting point for a branch-and-bound algorithm analogous to that of the three-way design.

## 4.5 Conclusions and Discussion

The results from this and other studies (Sanders, Theunissen & Baas 1989, 1991) show how integer optimization methods can be applied to find the optimal design for the maximization of a ratio of variance components. In general, these methods are ultimately suitable for solving problems for which analytical solutions are hard to obtain. One of these problems - already discussed by Scheffé in 1959 (pp. 236-238) - is to find the optimal design for the estimation of (ratios of) variance components. It was only recently, however, that Mukerjee and Huda (1988, pp. 78-79) suggested the use of enumeration methods for solving this problem under restrictions on total cost and other practical constraints. In this article a procedure is presented which can be used to solve some specific design problems.

## References


Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Graybill, F.A. & Wang, C.M. (1979). Confidence intervals for proportions of variability in two-factor nested variance component models. Journal of the American Statistical Association 74, 368-374.

Mukerjee, R. & Huda, S. (1988). Optimal design for the estimation of variance components. *Biometrika 75*, 75-80.

Papadimitriou, Ch.H. & Steiglitz, K. (1982). *Combinatorial optimization: Algorithms and complexity*. Englewood Cliffs: Prentice-Hall.

Sanders, P.F., Theunissen, T.J.J.M. & Baas, S.M. (1989). Minimizing the number of observations: A generalization of the Spearman-Brown formula. *Psychometrika 54*, 587-598.

Sanders, P.F., Theunissen, T.J.J.M. & Baas, S.M. (1991). Maximizing the coefficient of generalizability under the constraint of limited resources. *Psychometrika 56*, 87-96.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

Woodward, J.A. & Joe, G.W. (1973). Maximizing the coefficient of generalizability in multi-facet decision studies. *Psychometrika 38*, 173-181.

# Chapter 5

# Alternative Solutions for Optimization Problems in Generalizability Theory*

* Most of the material presented in this chapter will be published in Sanders, P.F. (in press). Alternative solutions for optimization problems in generalizability theory. *Psychometrika*.

## 5.1 Introduction

A procedure for the problem of maximizing the generalizability coefficient of the two-facet random-model crossed design under a budget constraint was recently presented by Sanders, Theunissen, and Baas (1991), who formulated this optimization problem as:

$$\text{minimize} \qquad \frac{\sigma^2_{p1}}{n_1} + \frac{\sigma^2_{p2}}{n_2} + \frac{\sigma^2_{res}}{n_1 n_2} = v, \tag{1}$$

$$\text{subject to} \qquad c_1 n_1 + c_2 n_2 + c_{12} n_1 n_2 \leq C, \text{ and} \tag{2}$$

$$n_1 \text{ and } n_2 \text{ integer} \geq 1. \tag{3}$$

Maximizing the generalizability coefficient is equivalent to minimizing the error variance, that is, the sum of the three interaction variance components. In (1), $\sigma^2_{p1}$ is the variance component for the person by Facet 1 interaction, $\sigma^2_{p2}$ is the variance component for the person by Facet 2 interaction, $\sigma^2_{res}$ is the variance component for the person x Facet 1 x Facet 2 interaction plus error, and $n_1$ and $n_2$ are the number of conditions of Facet 1 and Facet 2. The minimization statement in (1) expresses that the value of the objective-function is determined by the numbers of conditions, $n_1$ and $n_2$, used for Facet 1 and 2.

In the cost function in (2), the right-hand term specifies an upper limit on the budget. The costs required by the conditions of Facet 1 are specified by the term $c_1 n_1$, $c_1$ being the cost of one condition of Facet 1. The costs required by the conditions of Facet 2 are specified by the term $c_2 n_2$, $c_2$ being the cost of one condition of Facet 2. The number of observations per subject equals $n_1 n_2$. Thus, denoting the cost of one observation for the sample of subjects to be tested by $c_{12}$, the costs necessary for the total number of observations are specified by the term $c_{12} n_1 n_2$.

The lower bound integer constraint (3) states that feasible values for $n_1$ and $n_2$ have to be integer values and each facet has to have at least one condition.

In the two-step procedure proposed by Sanders et al. (1991), optimal continuous solutions for $n_1$ and $n_2$ are derived for a continuous relaxation of (3) in the first step. These continuous solutions are then used as the bounds in a branch-and-bound algorithm to obtain the optimal integer solutions. The use of the Lagrange multipliers method to attain the optimal continuous solutions, however, can result in complex derivations. But, as first shown by Stuart (1954) for problems of sample survey theory, derivations of optimum solutions can be simplified if the Cauchy-Schwarz inequality applies. This inequality states that for any sets of real numbers

$\{a_h\}$, $\{b_h\}$, $(h = 1, 2, \ldots, p)$,

$$(\Sigma a_h^2)(\Sigma b_h^2) \geq (\Sigma a_h b_h)^2, \tag{4}$$

with equality occurring if and only if $a_h/b_h = k$ for all $h$ and some constant $k$. According to Stuart (1954, p. 239), the sampling variance in most of the problems in sample survey theory takes the form $\Sigma \frac{v_h}{n_h} = v$, where $v_h$ is the function of population parameters only, and $n_h$ is a function of sample numbers only. The cost functions generally considered are of the form $\Sigma n_h c_h = c$, where the $c_h$ are fixed cost constants. From (4) it therefore follows that

$$vc = \left(\Sigma \frac{v_h}{n_h}\right)(\Sigma n_h c_h) \geq \left(\Sigma (v_h c_h)^{1/2}\right)^2. \tag{5}$$

Since the right-hand side of (5) is independent of the $n_h$, the minimization of $vc$ for fixed $c$ (or for fixed $v$) and variation in the $n_h$ is achieved when, from (4) and (5),

$$\frac{n_h^2 c_h}{v_h} = k^2 \text{ (all } h). \tag{6}$$

In the survey sampling literature (e.g., Cochran, 1977), applications of the Cauchy-Schwarz inequality to optimization problems

abound. The purpose of the present note is to show its applicability to optimization problems occurring in generalizability theory.

## 5.2 Solutions for Two-Facet Optimization Problems

Other linear cost functions besides (2) can arise if $C_1$, $C_2$ or $C_{12}$ is equal to zero. Functions with $C_1$ or $C_2$ and $C_{12}$ equal to zero, however, lead to trivial optimization problems. Moreover, the cost function with $C_1$ and $C_2$ equal to zero, that is, cost function $\frac{C}{C_{12}} \leq n_1 n_2$, reduces the problem of optimum allocation for fixed cost to the problem of optimum allocation for a fixed number of observations. By replacing the inequality constraint in the cost function by an equality constraint, the optimal continuous solutions for this optimization problem,

$$n_1 = \left( \frac{\sigma_{p1}^2}{\sigma_{p2}^2} \cdot \frac{C}{C_{12}} \right)^{1/4}, \text{ and } n_2 = \left( \frac{\sigma_{p2}^2}{\sigma_{p1}^2} \cdot \frac{C}{C_{12}} \right)^{1/4},$$

were derived by Woodward and Joe (1973). Therefore, three other potential linear cost functions can be distinguished:

(7) $C_1 n_1 + C_{12} n_1 n_2 \leq C$,

(8) $C_2 n_2 + C_{12} n_1 n_2 \leq C$, and

(9) $C_1 n_1 + C_2 n_2 \leq C$.

These three cost functions and (1) result in three different optimization problems.

The solution strategy for each of the three optimization problems is as follows. First, the inequality constraint of the cost function is replaced by an equality constraint. Second, the terms in the objective-function and the cost function are reformulated so that (5) applies. Hereafter we will illustrate the solution strategy of the optimization problem defined by (1) and (2) in detail and summarize the solutions for cost functions (7), (8), and (9). A detailed derivation of the latter solutions is presented in Sanders (1991).

We start by dividing (2) by $n_1 n_2$ which results in $\frac{c_1}{n_2} + \frac{c_2}{n_1} - \frac{c}{n_1 n_2} = -c_{12}$. By adding a constant to (1) and multiplying this constant by $-c_{12}$, we obtain

$$\frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2} + \frac{\sigma_{res}^2}{n_1 n_2} + \frac{\sigma_{res}^2}{c} \cdot -c_{12} = \frac{\left(\sigma_{p1}^2 + \frac{\sigma_{res}^2 c_2}{c}\right)}{n_1} + \frac{\left(\sigma_{p2}^2 + \frac{\sigma_{res}^2 c_1}{c}\right)}{n_2} \qquad (10)$$

Adding $\frac{c_1 c_2}{c_{12}}$ to both sides of (2) results in

$$c_1\left(n_1 + \frac{c_2}{c_{12}}\right) + c_{12}\left(n_1 + \frac{c_2}{c_{12}}\right)n_2 = \left(c + \frac{c_1 c_2}{c_{12}}\right) \rightarrow$$

$$c_{12} = -\frac{c_1}{n_2} + \frac{\left(c + \frac{c_1 c_2}{c_{12}}\right)}{\left(n_1 + \frac{c_2}{c_{12}}\right)n_2}.$$

By adding a constant to (10) and multiplying this constant by $c_{12}$ we obtain

$$\frac{\left(\sigma_{p1}^2 + \frac{\sigma_{res}^2 c_2}{c}\right)}{n_1} + \frac{\left(\sigma_{p2}^2 + \frac{\sigma_{res}^2 c_1}{c}\right)}{n_2} + \frac{\left(\sigma_{p2}^2 + \frac{\sigma_{res}^2 c_1}{c}\right)}{c_1} \cdot c_{12} =$$

$$\frac{\left(\sigma_{p1}^2 + \frac{\sigma_{res}^2 c_2}{c}\right)}{n_1} + \frac{\left(c + \frac{c_1 c_2}{c_{12}}\right)\left(\frac{\sigma_{p2}^2 c + \sigma_{res}^2 c_1}{c c_1}\right)}{\left(n_1 + \frac{c_2}{c_{12}}\right)n_2}. \qquad (11)$$

As cost function (2) can also be stated as

$$c_1 n_1 + c_{12}\left(n_1 + \frac{c_2}{c_{12}}\right)n_2 = c, \qquad (12)$$

we obtain by applying the Cauchy-Schwarz inequality to (11) and (12)

$$\frac{\left(\sigma_{p1}^2 + \frac{\sigma_{res}^2 c_2}{c}\right)}{n_1} \Big/ c_1 n_1 = k^2, \text{ and} \qquad (13)$$

$$\frac{\left(c+\frac{c_1c_2}{c_{12}}\right)\left(\frac{\sigma_{p2}^2c+\sigma_{res}^2c_1}{cc_1}\right)}{\left(n_1+\frac{c_2}{c_{12}}\right)n_2}\Bigg/\, c_{12}\left(n_1+\frac{c_2}{c_{12}}\right)n_2 = k^2. \tag{14}$$

Taking the square root of (13) and (14) and multiplying (13) by $(c_1)^{1/2}$ and (14) by $(c_{12})^{1/2}$ results in

$$\left(\sigma_{p1}^2+\frac{\sigma_{res}^2c_2}{c}\right)^{1/2}(c_1)^{1/2} = k\,c_1n_1, \text{ and} \tag{15}$$

$$\left(c+\frac{c_1c_2}{c_{12}}\right)^{1/2}\left(\frac{\sigma_{p2}^2c+\sigma_{res}^2c_1}{cc_1}\right)^{1/2}(c_{12})^{1/2} = k\,c_{12}\left(n_1+\frac{c_2}{c_{12}}\right)n_2, \tag{16}$$

from which it follows that

$$k = \frac{\left(\sigma_{p1}^2+\frac{\sigma_{res}^2c_2}{c}\right)^{1/2}(c_1)^{1/2}+\left(c+\frac{c_1c_2}{c_{12}}\right)^{1/2}\left(\frac{\sigma_{p2}^2c+c_1\sigma_{res}^2}{cc_1}\right)^{1/2}(c_{12})^{1/2}}{c}.$$

Having determined $k$, $n_1$ and $n_2$ can be solved.

For the optimization problem defined by (1) and (7), both functions need to be rewritten in order to apply the Cauchy-Schwarz inequality. After some algebraic manipulations we obtain

$$k = \frac{(\sigma_{p1}^2)^{1/2}(c_1)^{1/2}+\left(\sigma_{res}^2+\sigma_{p2}^2\cdot\frac{c}{c_1}\right)^{1/2}(c_{12})^{1/2}}{c}.$$

Having determined $k$, $n_1$ and $n_2$ can be solved. Since the functions of the optimization problem defined by (1) and (8) are equivalent to the functions defined by (1) and (7), the derivation of the solution for this problem is analogous.

In order to apply the Cauchy-Schwarz inequality to the optimization problem defined by (1) and (9), also requires rewriting both functions. After some algebraic maipulations we obtain

$$k = \frac{\left(\sigma^2_{p1} + \frac{\sigma^2_{res}C_2}{C}\right)^{1/2}(C_1)^{1/2} + \left(\sigma^2_{p2} + \frac{\sigma^2_{res}C_1}{C}\right)^{1/2}(C_2)^{1/2}}{C}.$$

Having determined $k$, $n_1$ and $n_2$ can be solved.

### 5.3 Two-Facet Example

Using variance components $\hat{\sigma}^2_p = 5.435$, $\hat{\sigma}^2_{p1} = 3.421$, $\hat{\sigma}^2_{p2} = 1.140$, and $\hat{\sigma}^2_{res} = 11.850$, the objective-function is defined as:

minimize $\quad \frac{3.421}{n_1} + \frac{1.140}{n_2} + \frac{11.850}{n_1 n_2}$ .

Let $C_1$, $C_2$ and $C_{12}$, of cost functions (2), (7), (8), and (9) be the cost of a condition of Facet 1 (e.g., essay questions), the cost of a condition of Facet 2 (for example, raters), and the cost of one observation for all subjects in the sample (i.e., the answers of all students to one essay question rated by one rater), respectively. With a total budget of 3000 dollars, the following three cost functions, leading to three optimization problems, are considered: $40n_1 + 80n_1n_2 \leq 3000$, $30n_1 + 80n_2 \leq 3000$, and $40n_1 + 80n_2 + 80n_1n_2 \leq 3000$. The optimal integer solutions $\hat{n}_1$ and $\hat{n}_2$ for the three optimization problems are derived in two steps. First, the optimal continuous solutions are determined with the equations presented before. These solutions are $n_1^* = 8.8$ and $n_2^* = 3.8$ for the first problem, $n_1^* = 40.9$ and $n_2^* = 17$ for the second problem, and $n_1^* = 9.1$ and $n_2^* = 3.3$ for the third problem. Second, a branch-and-bound procedure such as described in Sanders et al. (1991) is used to obtain the optimal integer solutions. Table 1 contains the results for the three problems.

**TABLE 1**

Values of $n_1$, $n_2$, Variance Components, $\rho^2$ and $C$ for Three Cost Functions.

| $n_1$ | $c_1 n_1$ | $n_2$ | $c_2 n_2$ | $n_1 n_2$ | $c_{12} n_1 n_2$ | $\hat{\sigma}_p^2$ | $\frac{\hat{\sigma}_{p1}^2}{n_1}$ | $\frac{\hat{\sigma}_{p2}^2}{n_2}$ | $\frac{\hat{\sigma}_{res}^2}{n_1 n_2}$ | $\rho^2$ | $C$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. $C_1$ = 40 dollars, $C_2$ = 0 dollars, and $C_{12}$ = 80 dollars | | | | | | | | | | | |
| 8.8 | 351 | 3.8 | 0 | 33.1 | 2649 | 5.435 | .390 | .302 | .358 | .838 | 3000 |
| 6 | 240 | 5 | 0 | 30 | 2400 | 5.435 | .570 | .228 | .395 | .820 | 2640 |
| 7 | 280 | 5 | 0 | 35 | 2800 | 5.435 | .489 | .228 | .339 | .837 | 3080 |
| 8* | 320 | 4* | 0 | 32 | 2560 | 5.435 | .428 | .285 | .370 | .834 | 2880 |
| 9 | 360 | 4 | 0 | 36 | 2880 | 5.435 | .380 | .285 | .329 | .845 | 3240 |
| 10 | 400 | 3 | 0 | 30 | 2400 | 5.435 | .342 | .380 | .395 | .830 | 2800 |
| 11 | 440 | 3 | 0 | 33 | 2640 | 5.435 | .311 | .380 | .359 | .838 | 3080 |
| 2. $C_1$ = 40 dollars, $C_2$ = 80 dollars, and $C_{12}$ = 0 dollars | | | | | | | | | | | |
| 40.9 | 1635 | 17 | 1364 | 697.5 | 0 | 5.435 | .084 | .067 | .017 | .970 | 3000 |
| 39* | 1560 | 18* | 1440 | 702 | 0 | 5.435 | .088 | .063 | .017 | .970 | 3000 |
| 40 | 1600 | 17 | 1360 | 680 | 0 | 5.435 | .086 | .067 | .017 | .970 | 2960 |
| 41* | 1640 | 17* | 1360 | 697 | 0 | 5.435 | .083 | .067 | .017 | .970 | 3000 |
| 3. $C_1$ = 40 dollars, $C_2$ = 80 dollars, and $C_{12}$ = 80 dollars | | | | | | | | | | | |
| 9.1 | 361 | 3.3 | 261 | 29.7 | 2375 | 5.435 | .376 | .350 | .399 | .829 | 3000 |
| 7* | 280 | 4* | 320 | 28 | 2240 | 5.435 | .489 | .285 | .432 | .819 | 2840 |
| 8 | 320 | 4 | 320 | 32 | 2560 | 5.435 | .428 | .285 | .370 | .834 | 3200 |
| 9 | 360 | 3 | 240 | 27 | 2160 | 5.435 | .380 | .380 | .439 | .819 | 2760 |
| 10 | 400 | 3 | 240 | 30 | 2400 | 5.435 | .342 | .380 | .395 | .830 | 3040 |

*Optimal integer solutions

## 5.4 Conclusions and Discussion

With respect to cost functions, the influence of the cost factors on the optimal solution should be noted. From the solutions in Table 1 it can be seen that the cost functions discussed here are dominated by cost factor $C_{12}$ = 80 dollars. The solutions for the first problem with cost function (7) and the third problem with cost function (2) are therefore very close, while they are quite different from the solutions for the second problem with cost function (9). These results, however, also imply that an optimization problem with a cost function including a dominant cost factor will give good starting solutions for a branch-

and-bound procedure for other optimization problems with cost functions including this dominant cost factor.

In survey sampling, (7) or (8) and (9) are the cost functions employed in connection with the allocation of optimal resources for designs known as two-stage and two-phase sampling designs. Applications of the Cauchy-Schwarz inequality to more complex designs, for example, three-stage sampling designs (e.g., Snedecor & Cochran 1976, p. 533) have been presented. As this note showed the formal similarity between optimization problems in survey sampling and generalizability theory, these applications signify that solutions for more complex designs employed in generalizability theory could also be derived.

## References


Cochran, W.G. (1977). *Sampling techniques*. New York: Wiley.

Sanders, P.F. (1991). *Alternative solutions for optimization problems in generalizability theory*. Measurement and Research Department Reports (91-9). Arnhem: Cito.

Sanders, P.F., Theunissen, T.J.J.M., & Baas, S.M. (1991). Maximizing the coefficient of generalizability under the constraint of limited resources. *Psychometrika, 56*, 87-96.

Snedecor, G.W., & Cochran, W.G. (1976). *Statistical methods*. Ames: Iowa State University Press.

Stuart, A. (1954). A simple presentation of optimum sampling results. *Journal of the Royal Statistical Society, B16*, 239-241.

Woodward, J.A., & Joe, G.W. (1973). Maximizing the coefficient of generalizability in multi-facet decision studies. *Psychometrika, 38*, 173-181.

# Chapter 6

# Sampling Errors of Variance Components

## 6.1 Introduction

Generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) makes a distinction between a generalizability study and a decision study. A generalizability study is conducted to obtain estimates of variance components associated with the universe of admissible observations. In a decision study these estimates are used to make decisions on the actual composition of a measurement instrument. Owing to limited resources, however, generalizability studies are often carried out with relatively small samples and relatively few conditions of each facet. Cronbach et al. (1972) have pointed out the danger of employing estimates of variance components and coefficients of generalizability studies based on small samples. The estimates of the variance components in these studies frequently entail sampling errors too large to be useful for a decision study.

A major study on sampling errors of variance components was conducted within the framework of generalizability theory by Smith (1978). The study employed an intuitive approach for solving the problem of how to allocate the number of conditions to different facets in order to produce the most stable estimate of the universe score variance. In foregoing chapters, the versatility of optimization techniques for solving allocation problems in decision studies was amply demonstrated. In this chapter, optimization techniques are proposed as a promising approach for solving allocation problems in generalizability studies.

Sampling errors of estimates of variance components are derived in section 6.2. The minimization of sampling errors of variance components under constraints is discussed in section 6.3. The balanced two-facet random-model crossed design will be used for illustration in the two sections.

## 6.2 Sampling Errors of Estimates of Variance Components

In generalizability theory, the analysis of variance method (e.g., Searle, 1971, p. 384) is used for the estimation of variance components from balanced data. This method starts with calculating the mean squares of the analysis of variance, whereafter the expected values of the mean squares are derived under the random or mixed model. Table 1 contains the expected mean squares for the two-facet random-model crossed design.

**TABLE 1**

Expected Mean Squares for the $p$ x $i$ x $j$ Random-Model Crossed Design

| Source of Variance | Expected Mean Squares |
|---|---|
| Persons($p$) | $EMS_p = \sigma^2_{pij,e} + n_i\sigma^2_{pj} + n_j\sigma^2_{pi} + n_in_j\sigma^2_p$ |
| Items($i$) | $EMS_i = \sigma^2_{pij,e} + n_p\sigma^2_{ij} + n_j\sigma^2_{pi} + n_pn_j\sigma^2_i$ |
| Raters($j$) | $EMS_j = \sigma^2_{pij,e} + n_p\sigma^2_{ij} + n_i\sigma^2_{pj} + n_pn_i\sigma^2_j$ |
| Persons x items($pi$) | $EMS_{pi} = \sigma^2_{pij,e} + n_j\sigma^2_{pi}$ |
| Persons x raters($pj$) | $EMS_{pj} = \sigma^2_{pij,e} + n_i\sigma^2_{pj}$ |
| Items x raters($ij$) | $EMS_{ij} = \sigma^2_{pij,e} + n_p\sigma^2_{ij}$ |
| Residual($pij,e$) | $EMS_{res} = \sigma^2_{pij,e}$ |

Equating the expected values to the calculated values leads to linear equations in the variance components, the solutions to which are the estimators of the variance components. The variance components for the two-facet random-model crossed design can be obtained with the following equations:

$$\hat{\sigma}^2_{pij,e} = MS_{pij,e}$$
$$\hat{\sigma}^2_{ij} = (MS_{ij} - MS_{pij,e})/n_p$$
$$\hat{\sigma}^2_{pj} = (MS_{pj} - MS_{pij,e})/n_i$$
$$\hat{\sigma}^2_{pi} = (MS_{pi} - MS_{pij,e})/n_j$$
$$\hat{\sigma}^2_j = (MS_j - MS_{ij} - MS_{pj} + MS_{pij,e})/(n_pn_i)$$
$$\hat{\sigma}^2_i = (MS_i - MS_{ij} - MS_{pi} + MS_{pij,e})/(n_pn_j)$$
$$\hat{\sigma}^2_p = (MS_p - MS_{pj} - MS_{pi} + MS_{pij,e})/(n_in_j)$$

In the statistical literature (e.g., Searle, 1971, pp. 415-417), the theoretical sampling variance of estimates of variance components is expressed in terms of the expected mean squares. As was shown in Table 1, the expected mean squares are linear combinations of the population values of the variance components, and therefore the expected sampling variance of the variance components of the two-facet random-model crossed design can be expressed as:

$$var\ \hat{\sigma}^2_p = \frac{2}{(n_p-1)}\left[\left(\sigma^2_p + \frac{\sigma^2_{pj}}{n_j} + \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{res}}{n_i n_j}\right)^2 + \frac{1}{(n_i-1)}\left(\frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{res}}{n_i n_j}\right)^2 \right. \tag{1a}$$

$$\left. + \frac{1}{(n_j-1)}\left(\frac{\sigma^2_{pj}}{n_j} + \frac{\sigma^2_{res}}{n_i n_j}\right)^2 + \frac{1}{(n_i-1)(n_j-1)}\left(\frac{\sigma^2_{res}}{n_i n_j}\right)^2\right],$$

$$var\ \hat{\sigma}^2_i = \frac{2}{(n_i-1)}\left[\left(\sigma^2_i + \frac{\sigma^2_{pi}}{n_p} + \frac{\sigma^2_{ij}}{n_j} + \frac{\sigma^2_{res}}{n_p n_j}\right)^2 + \frac{1}{(n_p-1)}\left(\frac{\sigma^2_{pi}}{n_p} + \frac{\sigma^2_{res}}{n_p n_j}\right)^2 \right. \tag{1b}$$

$$\left. + \frac{1}{(n_j-1)}\left(\frac{\sigma^2_{ij}}{n_j} + \frac{\sigma^2_{res}}{n_p n_j}\right)^2 + \frac{1}{(n_p-1)(n_j-1)}\left(\frac{\sigma^2_{res}}{n_p n_j}\right)^2\right],$$

$$var\ \hat{\sigma}^2_j = \frac{2}{(n_j-1)}\left[\left(\sigma^2_j + \frac{\sigma^2_{pj}}{n_p} + \frac{\sigma^2_{ij}}{n_i} + \frac{\sigma^2_{res}}{n_p n_i}\right)^2 + \frac{1}{(n_p-1)}\left(\frac{\sigma^2_{pj}}{n_p} + \frac{\sigma^2_{res}}{n_p n_i}\right)^2 \right. \tag{1c}$$

$$\left. + \frac{1}{(n_i-1)}\left(\frac{\sigma^2_{ij}}{n_i} + \frac{\sigma^2_{res}}{n_p n_i}\right)^2 + \frac{1}{(n_p-1)(n_i-1)}\left(\frac{\sigma^2_{res}}{n_p n_i}\right)^2\right],$$

$$var\ \hat{\sigma}^2_{pj} = \frac{2}{(n_p-1)(n_j-1)}\left[\left(\sigma^2_{pj} + \frac{\sigma^2_{res}}{n_i}\right)^2 + \frac{1}{(n_i-1)}\left(\frac{\sigma^2_{res}}{n_i}\right)^2\right], \tag{1d}$$

$$var\ \hat{\sigma}^2_{pi} = \frac{2}{(n_p-1)(n_i-1)}\left[\left(\sigma^2_{pi} + \frac{\sigma^2_{res}}{n_j}\right)^2 + \frac{1}{(n_j-1)}\left(\frac{\sigma^2_{res}}{n_j}\right)^2\right], \tag{1e}$$

$$var\ \hat{\sigma}^2_{ij} = \frac{2}{(n_i-1)(n_j-1)}\left[\left(\sigma^2_{ij} + \frac{\sigma^2_{res}}{n_p}\right)^2 + \frac{1}{(n_p-1)}\left(\frac{\sigma^2_{res}}{n_p}\right)^2\right], \tag{1f}$$

$$var\ \hat{\sigma}^2_{res} = \frac{2}{(n_p-1)(n_j-1)(n_i-1)}\sigma^4_{res}. \tag{1g}$$

The three factors affecting the magnitude of the sampling errors of the variance components can be inferred from equations (1a-1g). The first factor concerns the relative magnitudes of the population values of variance components. While the population values of the variance components are not known, equations (1a-1g) can be very useful for a researcher planning a generalizability study if prior information on the relative magnitude of the population values of variance components is available from comparable studies. In Table 2, for example, total variance is proportioned in a way that resembles estimates commonly observed in applications of the two-facet design with respect to judgmental data such as observations or ratings (Smith, 1976, p. 90). The residual component is represented by a small ($\sigma^2_{res} = 20$) and a large ($\sigma^2_{res} = 76$) error component.

**TABLE 2**

Variance Components of a Two-Facet Random-Model Crossed Design

| Source of Variance | Variance Components | Proportion of total variance | |
|---|---|---|---|
| | | small error | large error |
| Persons($p$) | $\sigma^2_p = 20$ | .09 | .07 |
| Items($i$) | $\sigma^2_i = 47$ | .21 | .17 |
| Raters($j$) | $\sigma^2_j = 12$ | .05 | .04 |
| Persons x items($pi$) | $\sigma^2_{pi} = 64$ | .29 | .23 |
| Persons x raters($pj$) | $\sigma^2_{pj} = 35$ | .16 | .13 |
| Items x raters($ij$) | $\sigma^2_{ij} = 23$ | .10 | .08 |
| Residual($pij,e$) | $\sigma^2_{res} = 20,76$ | .09 | .27 |

The effect of the second factor, design configuration, on the magnitude of the sampling error of the universe score variance can be seen by comparing equation (1a) to equation (2) for a design which has facet $J$ nested within facet $I$:

$$var\ \hat{\sigma}^2_p = \frac{2}{(n_p-1)}\left[\left(\sigma^2_p + \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{res}}{n_j n_i}\right)^2 + \frac{1}{(n_i-1)}\left(\frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{res}}{n_j n_i}\right)^2\right]. \qquad (2)$$

The comparison of the two equations indicates that the sampling error of the estimate of the universe score variance can be decreased by employing a nested instead of a crossed design.

The third factor concerns the number of measurement objects and the number of conditions of each facet. It is obvious that a researcher can avoid large sampling errors by employing large samples and large numbers of observations in a generalizability study. According to Smith (1978, p.322), however, many studies use rather small sample sizes. In the designs examined in his study, the number of subjects ($n_p$) were represented at three magnitudes: low ($n_p$ = 25), medium ($n_p$ = 50), and high ($n_p$ = 100). The number of conditions of each of the two facets ($n_i$ and $n_j$) were represented at two, four, and eight conditions.

## 6.3 Minimizing Sampling Errors of Estimates of Variance Components

In section 6.2, three factors affecting the sampling errors of variance components were identified. However, a researcher cannot influence the first factor, and the second factor implies changing the design of the generalizability study. Therefore, given a certain design choice, a researcher can only manipulate the third factor. In this section, two approaches for controlling these sampling errors are presented. The results of the approach followed by Smith (1976, 1978) will be examined first. An approach for solving allocation problems based on mathematical programming is presented next.

### 6.3.1 Intuitive Approach

Under the constraints of a fixed number of $n_p$ subjects, and a fixed number of $L = n_i n_j$ observations, Smith (1978) sought the solution for the problem of how to allocate the observations in order to minimize the sampling variance of the universe score variance. Since, according to Smith (1978, p.328), simple solutions for this minimization problem could not be derived analytically, a more

intuitive approach was chosen. From the examination of equation (1a) he concluded that as the residual component increases, equal allocation will maximize the divisor of the residual component and the optimal solution tends toward $n_i = n_j$. From a study by Woodward and Joe (1973) he borrowed the conclusion that as the residual component decreases, the allocation to $n_i$ and $n_j$ is determined by the ratio of $\sigma^2_{pi}$ and $\sigma^2_{pj}$. In order to evaluate these two conclusions, the minimization problem is stated in terms of mathematical programming as

$$\text{minimize} \quad var\ \hat{\sigma}^2_p \qquad \text{objective-function} \qquad (3)$$

$$\text{subject to } L = n_i n_j, \text{ and}$$

$$n_p = \text{constant.} \qquad \text{equality constraints} \qquad (4)$$

Objective-function (3) can be composed of four parts:

$$\left(\sigma^2_p + \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{pj}}{n_j} + \frac{\sigma^2_{res}}{n_i n_j}\right)^2, \qquad (3a)$$

$$\frac{1}{(n_i-1)}\left(\frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{res}}{n_i n_j}\right)^2, \qquad (3b)$$

$$\frac{1}{(n_j-1)}\left(\frac{\sigma^2_{pj}}{n_j} + \frac{\sigma^2_{res}}{n_i n_j}\right)^2, \qquad (3c)$$

$$\frac{1}{(n_i-1)\,(n_j-1)}\left(\frac{\sigma^2_{res}}{n_i n_j}\right)^2. \qquad (3d)$$

From equations (3a - 3d) it can be inferred that (3a), the 'persons' part, determines the value of the objective-function almost completely. With the appropriate variance components from Table 2 with a small error component and $n_i = 8$ and $n_j = 4$, the contribution of (3a) to the value of the objective-function is 97%. With a large error component,

this contribution is 96%. For both error components, the contribution of (3d), the 'residual' part, is negligible. The optimal continuous solutions for the variance components from Table 2 with $L$ = 32 and $\sigma^2_{res}$ equal to 20 or 76, are computed as $n_i$ = 7.4 and $n_j$ = 4.3. The same solutions are obtained with $\sigma^2_{res}$ equal to zero. With an improbable residual component equal to 200, however, the optimal continuous solutions are only slightly different, $n_i$ = 7.3 and $n_j$ = 4.4. Only an infinitely large residual component would affect the value of the objective-function and result in a solution tending toward $n_i = n_j$. The conclusion therefore is that the optimal number of conditions for the two facets does not depend on the magnitude of the residual component. To a lesser degree the same conclusion applies to the influence of parts (3b) and (3c), the 'interaction' components, of equation (1a).

Woodward and Joe (1973) showed that for the problem of maximizing the generalizability coefficient subject to the constraint that the total number of observations is fixed, the optimal allocation of conditions should be made directly proportional to $\sigma^2_{pi}/\sigma^2_{pj}$. In Chapter 3, their problem was stated in terms of mathematical programming as

$$\text{minimize} \quad \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{pj}}{n_j} + \frac{\sigma^2_{res}}{n_i n_j} \qquad \text{objective-function} \qquad (5)$$

$$\text{subject to } L = n_i n_j. \qquad \text{equality constraint} \qquad (6)$$

Because of equality constraint (6), the residual component does not play a role in determining the optimal number of conditions. The optimal continuous solutions were shown to be

$$n_i = \left( L \frac{\sigma^2_{pi}}{\sigma^2_{pj}} \right)^{1/2}, \text{ and } n_j = \left( L \frac{\sigma^2_{pj}}{\sigma^2_{pi}} \right)^{1/2}.$$

Smith (1978, p. 329) fails to explain why the results obtained for the optimization problem defined by (5) and (6) also apply to a seemingly quite different optimization problem as defined by (3) and (4). The reason is that the dominant role of part (3a) in objective-function (3) makes the problem of minimizing objective-function (3) almost identical to the problem of minimizing objective-function (5). Note that by taking the square root of (3a) and leaving out the constant $\sigma_p^2$, equation (5) is obtained. The continuous solutions for the minimization problem defined by (3) and (4) will therefore hardly differ from those for the minimization problem defined by (5) and (6). The optimal continuous solutions for the optimization problem defined by (5) and (6) for the variance components from Table 2 are $n_i = 7.6$ and $n_j = 4.2$.

### 6.3.2 Mathematical Programming Approach

Mathematical programming can be used to solve the problem of controlling the sampling errors of variance components. As in the foregoing chapters, the allocation problem is viewed as the minimization of sampling errors under constraints regarding observations and as the minimization of observations under constraints regarding sampling errors.

From the discussion in section 6.3.1 the conclusion can be drawn that for most generalizability studies the continuous solutions of optimization problems defined by (3) and (4) will not be much different from the continuous solutions of optimization problems defined by (5) and (6). A procedure for solving the optimization problem defined by (5) and (6) was presented in Chapter 3. This procedure, which also allows for less restrictive and more realistic constraints than (6), can be applied to the optimization problem under consideration.

Smith (1978, p. 327) only discusses the problem of minimizing the sampling error of the universe score variance. Mathematical

programming, however, offers the possibility to solve optimization problems with multiple objective-functions. Williams (1985, pp. 24-26) describes three ways of dealing with multiple objective-functions. First, the optimization problem can be solved with each objective-function in turn. A satisfactory solution for the optimization problem can be selected from the solutions resulting from these different objective-functions. Second, objectives and constraints can be interchanged, that is, treating all but one objective-function as a constraint. Third, a linear combination of all the objective-functions can be taken and this multiple objective-function is being optimized. The optimization problem

$$\text{minimize} \quad w_p var\ \hat{\sigma}_p^2 + w_i var\ \hat{\sigma}_i^2 + w_j var\ \hat{\sigma}_j^2 \qquad \text{objective-function}$$

$$\text{subject to } L = n_i n_j, \text{ and}$$

$$n_p = \text{constant}, \qquad \text{equality constraints}$$

is a version of the optimization problem defined by (3) and (4) with (3) now being a linear combination of three objective-functions. This solution to the optimization problem requires giving relative weights, $w_p$, $w_i$, and $w_j$, to the three objective-functions. A less restrictive formulation of this multi-objective optimization problem is obtained by introducing an inequality constraint for the sample size, e.g., $n_p \leq 100$. In earlier chapters it was shown that inequality constraints have many advantages over equality constraints.

The other approach for controlling sampling errors of variance components consists of minimizing the number of observations and/or the number of measurement objects under constraints of specific sampling errors. A general formulation of this optimization problem with one objective-function is

minimize $n_p n_i n_j$ objective-function

subject to $var\ \hat{\sigma}_p^2 \leq e_p$, threshold constraint

where $e_p$ stands for the value of the sampling error of the universe score variance considered to be acceptable.

Assuming a fixed number of subjects, a restricted version of the foregoing optimization problem is

minimize $n_i n_j$ objective-function

subject to $var\ \hat{\sigma}_p^2 \leq e_p$ threshold constraint.

Assuming that the sampling error of the universe score variance is defined by equation (5), solutions for this optimization problem can be obtained by the optimization procedure presented in Chapter 2. That procedure could also be used to obtain starting solutions for optimization problems with more complex objective-functions and/or constraints.

An example of an optimization problem with multiple constraints is

minimize $n_p n_i n_j$ objective-function

subject to $var\ \hat{\sigma}_p^2 \leq e_p$, and
$var\ \hat{\sigma}_i^2 \leq e_i$, and
$var\ \hat{\sigma}_j^2 \leq e_j$, threshold constraints

where $e_p$, $e_i$, and $e_j$, stand for the values of three sampling errors of variance components. Because the size of the sample is often the factor responsible for the major part of the costs of a generalizability study, the optimization problem could be formulated as

minimize $n_i n_j$ objective-function

subject to $n_p \leq L$ threshold constraints

$var\ \hat{\sigma}_p^2 \leq e_p$, and

$var\ \hat{\sigma}_i^2 \leq e_i$, and

$var\ \hat{\sigma}_j^2 \leq e_j$,

where $L$ is the maximum sample size that the researcher can afford.

## 6.4 Conclusions and Discussion

In the study presented in this chapter it was shown how our understanding of factors determining the sampling errors of variance components can be improved by viewing the problem of the optimal allocation of observations as an optimization problem. This approach leads to the important conclusion that allocation problems in generalizability studies have much in common with allocation problems in decision studies. An important practical consequence of this conclusion could be a considerable reduction of the effort required to develop optimization procedures and computer programs.

Apart from the three factors discussed in section 6.2, another factor affecting the magnitude of the sampling errors of variance components is the complexity of the design (Smith, 1981). The effect of this factor on the stability of the estimates can be shown by comparing equation (1a) to the equation for a one-facet random-model crossed ($p$ x $i$) design:

$$var\ \hat{\sigma}_p^2 = \frac{2}{(n_p-1)}\left[\left(\sigma_p^2 + \frac{\sigma_{res}^2}{n_i}\right)^2 + \frac{1}{(n_i-1)}\left(\frac{\sigma_{res}^2}{n_i}\right)^2\right].$$

From the comparison it seems clear that the sampling error of the universe score variance component of the one-facet random-model will be smaller than the sampling error of the universe score variance of the

two-facet random-model. However, there is an essential difference between the impact of this factor and the other three factors. Design complexity implies generalizations to different universes which means that the variance components and their estimates have different interpretations.

The results of Smith's study (1978, p. 336) show that the sampling errors for the universe score variance are intolerably large for practical purposes unless $n_p n_i n_j$ is 800 or larger. Bell (1986) found the same results in a study on simultaneous confidence intervals of variance components. It should be clear to the researcher planning a generalizability study that the problem of sampling errors may make the results of the study unusable. Another problem is that of limited resources (time, money, etc.) which may render the study infeasible. What is needed therefore is a procedure which enables the researcher to design a generalizability study in line with stated limits on sampling errors as well as resources. The initial impetus to such a procedure was presented in this chapter.

## References


Bell, J.E. (1986). Simultaneous confidence intervals for the linear functions of expected mean squares used in generalizability theory. *Journal of Educational Statistics, 11,* 197-205.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Searle, S.R. (1971). *Linear Models.* New York: Wiley.

Smith, P.L. (1976). *Sampling errors of variance components in multifacet generalizability studies.* Unpublished doctoral dissertation. Urbana-Champaign: University of Illinois.

Smith, P.L. (1978). Sampling errors of variance components in small sample multifacet generalizability studies. *Journal of Educational Statistics, 2,* 319-346.

Smith, P.L. (1981). Gaining accuracy in generalizability theory: using multiple designs. *Journal of Educational Measurement, 18,* 147-154.

Williams, H.P. (1985). *Model building in mathematical programming.* New York: Wiley.

# Chapter 7

# Epilogue

The foreword to the book by Arthanari and Dodge (1981) on mathematical programming in statistics, starts with C. R. Rao saying: 'All statistical procedures are, in the ultimate analysis, solutions to suitably formulated optimization problems.' In Chapter 1 it was noted that problems in test theory have only recently been formulated as optimization problems. In the present study allocation problems in generalizability theory were formulated as optimization problems. In Chapters 2 to 5 it was shown that by formulating allocation problems in D-studies as optimization problems, these problems can be solved through the application of mathematical programming techniques. In Chapter 6 it was shown that formulations and solutions of some allocation problems pertinent to G-studies appeared to be quite similar to formulations and solutions of allocation problems found in D-studies.

The two optimization problems in D-studies that were investigated in the present study were the minimization of the number of observations and the maximization of the generalizability coefficient. The optimality criterion in both optimization problems was the 'cost' of the measurement procedure. Cronbach (1976, p. 200) noted that the original presentation of generalizability theory was incomplete without the concepts of costs and utility. Since decision theory (Cronbach & Gleser, 1965) does deal with these concepts, Cronbach (1976, p. 201) suggested hybridizing decision theory and generalizability theory. No attempts, however, have been undertaken to integrate the two theories. The purpose of the proposed hybridization was the design of measurement procedures which collect information efficiently. It seems that this purpose has been accomplished by hybridizing generalizability theory and mathematical programming.

Mathematical programming models have been found to be applicable to test construction procedures in the three test theories currently most used in test theory - item response theory, classical test theory,

and generalizability theory (cf. Suen, 1990). What makes these models particularly useful for constructors of tests and measurement procedures is the possibility to employ test specifications. These specifications can relate to the psychometric properties of a test as well as other internal test attributes (e.g., test composition) and external factors (e.g., resources). It will take some time, however, before a large-scale employment of mathematical programming techniques in actual test construction will be observed. One reason is that literature on mathematical programming techniques appropriate for test constructors is lacking. Moreover there is a lack of software. User friendly software specifically designed for the optimal selection of items calibrated with item response theory has only recently become available. While mathematical programming techniques for the optimal selection of items calibrated with classical test theory have been developed (e.g., Adema, 1990), specific software has yet to be designed. Regarding generalizability theory, the only software available for selecting the optimal number of conditions of facets 'calibrated' with generalizability theory, is the software described in Chapters 1 and 2.

Recently, asymptotic variance expressions of generalizability coefficients were derived by Schroeder and Hakstian (1990, p. 434). These expressions can be used to construct $(1 - \alpha)$ confidence intervals for generalizability coefficients. Using these intervals in the optimization problems of the present study will increase the number of conditions. Instead of advising the test constructor to employ eleven questions and three raters, the advice could be given to employ fourteen questions and three raters in order to achieve the desired generalizability coefficient with the desired confidence.

Optimization problems play an important role in the design of experiments. Arthanari and Dodge (1981, p. 289 ff.) distinguish four major areas of interest. One area, choosing the optimal design among

the available designs with respect to a given criterion, comprises the optimization problems of the present study. The discussions related to other areas, for example, the construction of designs with desired properties, indicate a bright perspective for future applications of mathematical programming to test theory. Within (multivariate) generalizability theory, mathematical programming techniques could probably successfully be applied to, for example, multilevel analysis (Fyans, 1983) and tests developed from tables of specifications (Jarjoura & Brennan, 1983). It is clear, therefore, that the last line of Arthanari and Dodge's book (1981, p. 404) also applies to generalizability theory: 'It does not seem to end here.'

## References


Adema, J.J. (1990). *Models and algorithms for the construction of achievement tests*. Enschede: FEBO.

Arthanari, T.S., & Dodge, Y. (1981). *Mathematical programming in statistics*. New York: Wiley.

Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions* (2nd. ed.). Champaign-Urbana: University of Illinois Press.

Cronbach, L.J. (1976). On the design of educational measures. In D.N.M. de Gruijter, & L.J.T. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 199-208). New York: Wiley.

Fyans, L.J. (1983). Multilevel analysis and cross-level inference of validity using generalizability theory. In L.J. Fyans (Ed.), *Generalizability theory: Inferences and practical applications* (pp. 49-66). San Francisco: Jossey-Bass.

Jarjoura, D., & Brennan, R.L. (1983). Multivariate generalizability models for tests developed from tables of specifications. In L.J. Fyans (Ed.), *Generalizability theory: Inferences and practical applications* (pp. 83-101). San Francisco: Jossey-Bass.

Schroeder, M.L., & Hakstian, A.R. (1990). Inferential procedures for multifaceted coefficients of generalizability. *Psychometrika, 55*, 429-447.

Suen, H.K. (1990). *Principles of test theories*. Hillsdale: Erlbaum.

## Samenvatting

# De optimalisering van decisiestudies in de generaliseerbaarheidstheorie

Mathematische programmeringstechnieken hebben het construeren van toetsen in de klassieke testtheorie en de itemresponstheorie wezenlijk veranderd. Met behulp van deze technieken kunnen, gegeven een verzameling items met bekende psychometrische en onderwijskundige eigenschappen, toetsen geconstrueerd worden die optimaal voldoen aan de door toetsconstructeurs gespecificeerde psychometrische en onderwijskundige eisen.

In deze studie wordt de toepassing van mathematische programmeringstechnieken in de generaliseerbaarheidstheorie onderzocht. In de generaliseerbaarheidstheorie wordt een onderscheid gemaakt tussen generaliseerbaarheidsstudies en decisiestudies. Het doel van een generaliseerbaarheidsstudie is het bepalen van de eigenschappen van een meetinstrument. Een voorbeeld van een generaliseerbaarheidsstudie is een studie waarbij studenten vragen beantwoorden die door beoordelaars beoordeeld worden. Het meetinstrument bestaat hier uit het facet vragen en het facet beoordelaars, terwijl de studenten de meetobjecten zijn. In een generaliseerbaarheidsstudie worden met behulp van een variantie-analyse schattingen van variantiecomponenten verkregen. Die variantiecomponenten kunnen in decisiestudies gebruikt worden om beslissingen over meetobjecten te nemen of om het meetinstrument definitief vorm te geven. Dit laatste betekent in ons voorbeeld dat de constructeur het aantal vragen en het aantal beoordelaars vast moet stellen.

Aan de hand van onder andere bovengenoemd voorbeeld werden twee optimaliseringsproblemen onderzocht. Het eerste probleem was om met zo weinig mogelijk vragen en beoordelaars (= doelfunctie) een bepaalde minimale waarde voor de generaliseerbaarheidscoëfficiënt (= voorwaarde) te verkrijgen. Dit optimaliseringsprobleem bleek opgelost te kunnen worden met behulp van een branch-en-bound-methode. Deze methode bestaat hieruit dat de oplossingsverzameling gesplitst wordt in disjuncte deelverzamelingen en dat op deze deelverzamelingen ondergrenzen voor de

waarde van de doelfunctie berekend worden. Met deze methode is het mogelijk om behalve de genoemde psychometrische voorwaarde en de voorwaarde dat de oplossing voor het aantal vragen en het aantal beoordelaars geheeltallig moet zijn, ook andere voorwaarden te specificeren. Als voorwaarde kan de toetsconstructeur bijvoorbeeld specificeren dat het meetinstrument uit meer vragen dan beoordelaars moet bestaan of dat het meetinstrument minstens tien vragen moet bevatten. De branch-en-bound-methode bleek behalve genoemde praktische eigenschappen ook belangrijke theoretische eigenschappen te hebben. Voor meetinstrumenten bestaande uit meerdere facetten bleek de methode een generalisatie te zijn van de Spearman-Brown formule uit de klassieke testtheorie.

Het tweede probleem dat onderzocht werd betrof het realiseren van een zo hoog mogelijke waarde voor de generaliseerbaarheidscoëfficiënt (= doelfunctie) zonder daarbij een bepaald budget (= voorwaarde) te overschrijden. Voor bovengenoemd voorbeeld bestond de voorwaarde uit een kostenfunctie bestaande uit kosten die veroorzaakt worden door vragen, beoordelaars en de beoordeling van de antwoorden van de studenten op de vragen. Dit optimaliseringsprobleem bleek in twee stappen opgelost te kunnen worden. In de eerste stap werd het probleem opgelost zonder dat als voorwaarde gesteld werd dat het aantal vragen en het aantal beoordelaars geheeltallig moest zijn. Een interessant theoretisch resultaat was dat de oplossing voor dit probleem ook met behulp van de Cauchy-Schwarz ongelijkheid afgeleid bleek te kunnen worden. In de tweede stap werden de optimale continue oplossingen gebruikt als grenzen voor een branch-en-bound-methode waarmee de optimale geheeltallige oplossingen voor dit optimaliseringsprobleem verkregen werden. De tweestaps procedure had als nadeel dat de afleiding van continue oplossingen in de eerste stap voor problemen bestaande uit meer dan twee facetten zeer complex kon worden. Voor eenvoudige kostenfuncties, bijvoorbeeld een kostenfunctie die slechts

bestaat uit kosten veroorzaakt door de beoordeling van de antwoorden, waren continue oplossingen voor problemen met meer dan twee facetten echter relatief eenvoudig af te leiden. Deze oplossingen bleken zeer goede startoplossingen voor een branch-en-bound-methode te zijn.

De toepassing van mathematische programmeringstechnieken bleek niet beperkt te zijn tot het optimaliseren van decisiestudies. Aangetoond werd dat het minimaliseren van standaardfouten van variantiecomponenten in generaliseerbaarheidsstudies ook als optimaliseringsprobleem opgevat kan worden. Tussen de oplossingen voor decisiestudies en generaliseerbaarheidsstudies werden opvallende overeenkomsten geconstateerd.

Geconcludeerd werd dat mathematische programmeringstechnieken ook bij andere problemen uit de generaliseerbaarheidstheorie toegepast zouden moeten worden.